CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04, C07K 14/00, C12N 15/00, C12P 19/34 — A1

(11) International Publication Number: **WO 98/42727**

(43) International Publication Date: 1 October 1998 (01.10.98)

(21) International Application Number: PCT/US98/05223

(22) International Filing Date: 16 March 1998 (16.03.98)

(30) Priority Data:

60/041,173	21 March 1997 (21.03.97)	US
08/910,367	13 August 1997 (13.08.97)	US

(71) Applicant: SRI INTERNATIONAL [US/US]; 333 Ravenswood Avenue, Menlo Park, CA 94025 (US).

(72) Inventors: PATI, Sushma; 1199 Cleveland Street, Redwood City, CA 94061 (US). ZARLING, David, A.; 1095 Colby Avenue, Menlo Park, CA 94025 (US).

(74) Agents: TRECARTIN, Richard, F. et al.; Flehr, Hohbach, Test, Albritton & Herbert LLP, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

(54) Title: SEQUENCE ALTERATIONS USING HOMOLOGOUS RECOMBINATION

(57) Abstract

The invention relates to methods for targeting an exogenous polynucleotide or exogenous complementary polynucleotide pair to a predetermined endogenous DNA target sequence in a target cell by homologous pairing, particularly for altering an endogenous DNA sequence, such as a chromosomal DNA sequence, typically by targeted homologous recombination. In certain embodiments, the invention relates to methods for targeting an exogenous polynucleotide having a linked chemical substituent to a predetermined endogenous DNA sequence in a metabolically active target cell, generating a DNA sequence-specific targeting of one or more chemical substituents in an intact nucleous of a metabolically active target cell, generally for purposes of altering a predetermined endogenous DNA sequence in the cell. The invention also relates to compositions that contain exogenous targeting polynucleotides, complementary pairs of exogenous targeting polynucleotides, chemical substituents of such polynucleotides, and recombinase proteins used in the methods of the invention.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SEQUENCE ALTERATIONS USING HOMOLOGOUS RECOMBINATION

FIELD OF THE INVENTION

The invention relates to methods for targeting an exogenous polynucleotide or exogenous complementary polynucleotide pair to a predetermined endogenous DNA target sequence in a target cell by homologous pairing, particularly for altering an endogenous DNA sequence, such as a chromosomal DNA sequence, typically by targeted homologous recombination. In certain embodiments, the invention relates to methods for targeting an exogenous polynucleotide having a linked chemical substituent to a predetermined endogenous DNA sequence in a metabolically active target cell, generating a DNA sequence-specific targeting of one or more chemical substituents in a metabolically active living target cell, generally for purposes of altering a predetermined endogenous DNA sequence in the cell. The invention also relates to compositions and formulations that contain exogenous targeting polynucleotides, complementary pairs of exogenous targeting polynucleotides, chemical substituents of such polynucleotides, and recombinase proteins, including recombinosome proteins and other targeting proteins, used in the methods of the invention.

BACKGROUND

Homologous recombination (or general recombination) is defined as the exchange of homologous segments anywhere along a length of two DNA molecules. An essential feature of general recombination is that the enzymes responsible for the recombination event can presumably use any pair of homologous sequences as substrates, although some types of sequence may be favored over others. Both genetic and cytological studies have indicated

that such a crossing-over process occurs between pairs of homologous chromosomes during meiosis in higher organisms.

Alternatively, in site-specific recombination, exchange occurs at a specific site, as in the integration of phage λ into the *E. coli* chromosome and the excision of λ DNA from it. Site-specific recombination involves specific inverted repeat sequences; e.g. the Cre-loxP and FLP-FRT systems. Within these sequences there is only a short stretch of homology necessary for the recombination event, but not sufficient for it. The enzymes involved in this event generally cannot recombine other pairs of homologous (or nonhomologous) sequences, but act specifically.

Although both site-specific recombination and homologous recombination are useful mechanisms for genetic engineering of DNA sequences, targeted homologous recombination provides a basis for targeting and altering essentially any desired sequence in a duplex DNA molecule, such as targeting a DNA sequence in a chromosome for replacement by another sequence. Site-specific recombination has been proposed as one method to integrate transfected DNA at chromosomal locations having specific recognition sites (O'Gorman et al. (1991) Science 251: 1351; Onouchi et al. (1991) Nucleic Acids Res. 19: 6373). Unfortunately, since this approach requires the presence of specific target sequences and recombinases, its utility for targeting recombination events at any particular chromosomal location is severely limited in comparison to targeted general recombination.

For these reasons and others, targeted homologous recombination has been proposed for treating human genetic diseases. Human genetic diseases include (1) classical human genetic diseases wherein a disease allele having a mutant genetic lesion is inherited from a parent (e.g., adenosine deaminase deficiency, sickle cell anemia, thalassemias), (2) complex genetic diseases like cancer, where the pathological state generally results from one or more specific inherited or acquired mutations, and (3) acquired genetic disease, such as an integrated provirus (e.g., hepatitis B virus).

Homologous recombination has also been used to create transgenic animals. Transgenic animals are organisms that contain stably integrated copies of genes or gene constructs derived from another species in the chromosome of the transgenic animal. These animals can be generated by introducing cloned DNA constructs of the foreign genes into totipotent cells by a variety of methods, including homologous recombination. Animals that develop from genetically altered totipotent cells contain the foreign gene in all somatic cells and also in germ-line cells if the foreign gene was integrated into the genome of the recipient cell before the first cell division. Currently methods for producing transgenics have been performed on totipotent embryonic stem cells (ES) and with fertilized zygotes. ES cells have an advantage in that large numbers of cells can be manipulated easily by homologous recombination in vitro before they are used to generate transgenics. Currently, however, only embryonic stem cells from mice have been characterized as contributing to the germ line. Alternatively, DNA can also be introduced into fertilized oocytes by micro-injection into pronuclei which are then transferred into the uterus of a pseudo-pregnant recipient animal to develop to term. However because current homologous recombination methods are inefficient and it is not logistically possible to manipulate large numbers of fertilized zygotes, transgenic animals produced by zygote microinjection are generally the result of random integration (not targeted) of the gene construct. A few cases of relatively inefficient homologous recombination in mouse fertilized zygotes have been reported, however these methods have been only been applied to a few specific target genes (Brinster et al. (1989) PNAS 86: 7087; Susulic et al. (1995) JBC 49: 29483; Zimmer and Gruss (1989) Nature 338: 150] and the general utility of homologous recombination in zygotes for any desired target gene has not been observed.

Commercial applications to produce transgenic animals by homologous recombination include 1) animal models to study gene function; 2) animal models that mimic human disease; 3) animals that produce therapeutic proteins from a known, pre-designated stable site in the chromosome; 4) animals that produce milk with superior nutritional value; 5) animal livestock with superior qualities, including disease and pathogen resistance; and 6) genetically altered animals that produce organs that are suitable for xenotransplantation. However as stated above, current methods for homologous recombination are generally inefficient and since ES cells which contribute to the germ line have only been identified for

mice, homologous recombination has not been enabled for producing transgenic animals in any other species other than two strains of mice.

Thus, current methods of targeted homologous recombination are inefficient and produce desired homologous recombinants only rarely, necessitating complex cell selection schemes to identify and isolate correctly targeted recombinants.

A primary step in homologous recombination is DNA strand exchange, which involves a pairing of a DNA duplex with at least one DNA strand containing a complementary sequence to form an intermediate recombination structure containing heteroduplex DNA (see, Radding, C.M. (1982) Ann. Rev. Genet. 16: 405; U.S. Patent 4,888,274). The heteroduplex DNA may take several forms, including a three DNA strand containing triplex form wherein a single complementary strand invades the DNA duplex (Hsieh et al. (1990) Genes and Development 4: 1951; Rao et al., (1991) PNAS 88:2984)) and, when two complementary DNA strands pair with a DNA duplex, a classical Holliday recombination joint or chi structure (Holliday, R. (1964) Genet. Res. 5: 282) may form, or a double-D loop ("Diagnostic Applications of Double-D Loop Formation" U.S.S.N. 07/755,462, filed 4 September 1991, which is incorporated herein by reference). Once formed, a heteroduplex structure may be resolved by strand breakage and exchange, so that all or a portion of an invading DNA strand is spliced into a recipient DNA duplex, adding or replacing a segment of the recipient DNA duplex. Alternatively, a heteroduplex structure may result in gene conversion, wherein a sequence of an invading strand is transferred to a recipient DNA duplex by repair of mismatched bases using the invading strand as a template (Genes, 3rd Ed. (1987) Lewin, B., John Wiley, New York, NY; Lopez et al. (1987) Nucleic Acids Res. 15: 5643). Whether by the mechanism of breakage and rejoining or by the mechanism(s) of gene conversion, formation of heteroduplex DNA at homologously paired joints can serve to transfer genetic sequence information from one DNA molecule to another.

The ability of homologous recombination (gene conversion and classical strand breakage/rejoining) to transfer genetic sequence information between DNA molecules makes

targeted homologous recombination a powerful method in genetic engineering and gene manipulation.

The ability of mammalian and human cells to incorporate exogenous genetic material into genes residing on chromosomes has demonstrated that these cells have the general enzymatic machinery for carrying out homologous recombination required between resident and introduced sequences. These targeted recombination events can be used to correct mutations at known sites, replace genes or gene segments with defective ones, or introduce foreign genes into cells. The efficiency of such gene targeting techniques is related to several parameters: the efficiency of DNA delivery into cells, the type of DNA packaging (if any) and the size and conformation of the incoming DNA, the length and position of regions homologous to the target site (all these parameters also likely affect the ability of the incoming homologous DNA sequences to survive intracellular nuclease attack), the efficiency of hybridization and recombination at particular chromosomal sites and whether recombinant events are homologous or nonhomologous. Over the past 10 years or so, several methods have been developed to introduce DNA into mammalian cells: direct needle microinjection, transfection, electroporation, retroviruses, adenoviruses, adeno-associated viruses; Herpes viruses, and other viral packaging and delivery systems, polyamidoamine dendimers, liposomes, and more recently techniques using DNA-coated microprojectiles delivered with a gene gun (called a biolistics device), or narrow-beam lasers (laser-poration). The processes associated with some types of gene transfer have been shown to be pathogenic, mutagenic or carcinogenic (Bardwell, (1989) Mutagenesis 4: 245), and these possibilities must be considered in choosing a transfection approach.

The choice of a particular DNA transfection procedure depends upon its availability to the researcher, the technique's efficiency with the particular chosen target cell type, and the researchers concerns about the potential for generating unwanted genome mutations. For example, retroviral integration requires dividing cells, most often results in nonhomologous recombination events, and retroviral insertion within a coding sequence of nonhomologous (i.e., non-targeted) gene could cause cell mutation by inactivating the gene's coding sequence (Friedmann, (1989) Science 244:1275). Newer retroviral-based DNA delivery systems are

being developed using modified retroviruses. However, these disabled viruses must be packaged using helper systems, are often obtained at low titer, and recombination is still not site-specific, thus recombination between endogenous cellular retrovirus sequences and disabled virus sequences could still produce wild-type retrovirus capable of causing gene mutation. Adeno- or polyoma virus based delivery systems appear promising (Samulski et al., (1991) EMBO J. 10: 2941; Gareis et al., (1991) Cell. Molec. Biol. 37: 191; Rosenfeld et al. (1992) Cell 68: 143) although they still require specific cell membrane recognition and binding characteristics for target cell entry. Liposomes often show a narrow spectrum of cell specificities, and when DNA is coated externally on to them, the DNA is often sensitive to cellular nucleases. Newer polycationic lipospermines compounds exhibit broad cell ranges (Behr et al., (1989) Proc. Natl. Acad. Sci. USA 86: 6982) and DNA is coated by these compounds. In addition, a combination of neutral and cationic lipid has been shown to be highly efficient at transfection of animal cells and showed a broad spectrum of effectiveness in a variety of cell lines (Rose et al., (1991) BioTechniques 10:520). Galactosylated bis-acridine has also been described as a carrier for delivery of polynucleotides to liver cells (Haensler JL and Szoka FC (1992), Abstract V211 in J. Cell. Biochem. Supplement 16F, April 3-16, 1992, incorporated herein by reference). Electroporation also appears to be applicable to most cell types. The efficiency of this procedure for a specific gene is variable and can range from about one event per 3 x 10^4 transfected cells (Thomas and Capecchi, (1987) Cell 51: 503) to between one in 10^7 and 10^8 cells receiving the exogenous DNA (Koller and Smithies, (1989) Proc. Natl. Acad. Sci. (U.S.A.) 86: 8932). Microinjection of exogenous DNA into the nucleus has been reported to result in stable integration in transfected cells. Zimmer and Gruss (Zimmer and Gruss (1989) Nature 338: 150) have reported that for the mouse hox1.1 gene, 1 per 150 microinjected cells showed a stable homologous site specific alteration.

Several methods have been developed to detect and/or select for targeted site-specific recombinants between vector DNA and the target homologous chromosomal sequence (see, Capecchi, (1989) Science 244: 1288 for review). Cells which exhibit a specific phenotype after site-specific recombination, such as occurs with alteration of the hprt gene, can be obtained by direct selection on the appropriate growth medium. Alternatively, a selective

marker sequence such as neo can be incorporated into a vector under promoter control, and successful transfection can be scored by selecting G418^r cells followed by PCR to determine whether neo is at the targeted site (Joyner et al., (1989) Nature 338: 153). A positive-negative selection (PNS) procedure using both neo and HSV-tk genes allows selection for transfectants and against nonhomologous recombination events, and significantly enriched for desired disruption events at several different mouse genes (Mansour et al., (1988) Nature 336: 348). This procedure has the advantage that the method does not require that the targeted gene be transcribed. If the targeted gene is transcribed, a promoter-less marker gene can be incorporated into the targeting construct so that the gene becomes activated after homologous recombination with the target site (Jasin and Berg, (1988) Genes and Development 2: 1353; Doetschman et al. (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 8583; Dorini et al., (1989) Science 243: 1357; Itzhaki and Porter, (1991) Nucl. Acids Res. 19: 3835). Recombinant products produced using vectors with selectable markers often continue to retain these markers as foreign genetic material at the site of transfection, although loss does occur. Valancius and Smithies (Valancius and Smithies, (1991) Mole. Cellular Biol. 11: 1402) have described an "in-out" targeting procedure that allowed a subtle 4-bp insertion modification of a mouse hprt target gene. The resulting transfectant contained only the desired modified gene sequence and no selectable marker remained after the "out" recombination step. Cotransformation of cells with two different vectors, one vector contained a selectable gene and the other used for gene disruption, increases the efficiency of isolating a specific targeting reaction (Reid et al., (1991) Molec. Cellular Biol. 11: 2769) among selected cells that are subsequently scored for stable recombinants.

Unfortunately, exogenous sequences transferred into eukaryotic cells undergo homologous recombination with homologous endogenous sequences only at very low frequencies, and are so inefficiently recombined that large numbers of cells must be transfected, selected, and screened in order to generate a desired correctly targeted homologous recombinant (Kucherlapati et al. (1984) Proc. Natl. Acad. Sci. (U.S.A.) 81: 3153; Smithies, 0. (1985) Nature 317: 230; Song et al. (1987) Proc. Natl. Acad. Sci. (U.S.A.) 84: 6820; Doetschman et al. (1987) Nature 330: 576; Kim and Smithies (1988) Nucleic Acids Res. 16: 8887; Doetschman et al. (1988) op.cit.; Koller and Smithies (1989) op.cit.; Shesely et al. (1991)

Proc. Natl. Acad. Sci. (U.S.A.) 88: 4294; Kim et al. (1991) Gene 103: 227, which are incorporated herein by reference).

Koller et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.), 88: 10730 and Snouwaert et al. (1992) Science 257: 1083, have described targeting of the mouse cystic fibrosis transmembrane regulator (CFTR) gene for the purpose of inactivating, rather than correcting, a murine CFTR allele. Koller et al. employed a large (7.8kb) homology region in the targeting construct, but nonetheless reported a low frequency for correct targeting (only 1 of 2500 G418-resistant cells were correctly targeted). Thus, even targeting constructs having long homology regions are inefficiently targeted.

Several proteins or purified extracts having the property of promoting homologous recombination (i.e., recombinase activity) have been identified in prokaryotes and eukaryotes (Cox and Lehman (1987) Ann. Rev. Biochem. 56: 229; Radding, C.M. (1982) op.cit.; Madiraju et al. (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 6592; McCarthy et al. (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 5854; Lopez et al. (1987) op.cit., which are incorporated herein by reference). These general recombinases presumably promote one or more steps in the formation of homologously-paired intermediates, strand-exchange, gene conversion, and/or other steps in the process of homologous recombination.

The frequency of homologous recombination in prokaryotes is significantly enhanced by the presence of recombinase activities. Several purified proteins catalyze homologous pairing and/or strand exchange in vitro, including: *E. coli* recA protein, the T4 uvsX protein, the rec1 protein from *Ustilago maydis*, and Rad51 protein from S. cervisiae (Sung et al., Science 265:1241 (1994)) and human cells (Baumann et al., Cell 87:757 (1996)). Recombinases, like the recA protein of *E. coli* are proteins which promote strand pairing and exchange. The most studied recombinase to date has been the recA recombinase of *E. coli*, which is involved in homology search and strand exchange reactions (see, Cox and Lehman (1987) op.cit.). RecA is required for induction of the SOS repair response, DNA repair, and efficient genetic recombination in *E. coli*. RecA can catalyze homologous pairing of a linear duplex DNA and a homologous single strand DNA in vitro. In contrast to site-specific recombinases, proteins

like recA which are involved in general recombination recognize and promote pairing of DNA structures on the basis of shared homology, as has been shown by several in vitro experiments (Hsieh and Camerini-Otero (1989) J. Biol. Chem. 264: 5089; Howard-Flanders et al. (1984) Nature 309: 215; Stasiak et al. (1984) Cold Spring Harbor Symp. Quant. Biol. 49: 561; Register et al. (1987) J. Biol. Chem. 262: 12812). Several investigators have used recA protein in vitro to promote homologously paired triplex DNA (Cheng et al. (1988) J. Biol. Chem. 263: 15110; Ferrin and Camerini-Otero (1991) Science 354: 1494; Ramdas et al. (1989) J. Biol Chem. 264: 11395; Strobel et al. (1991) Science 254: 1639; Hsieh et al. (1990) op.cit.; Rigas et al. (1986) Proc. Natl. Acad. Sci. (U.S.A.) 83: 9591; and Camerini-Otero et al. U.S. 7,611,268 (available from Derwent), which are incorporated herein by reference). Unfortunately many important genetic engineering manipulations involving homologous recombination, such as using homologous recombination to alter endogenous DNA sequences in a living cell, cannot be done in vitro. Further, gene therapy and transgenesis requires highly efficient homologous recombination of targeting vectors with predetermined endogenous target sequences, since selectable marker selection schemes such as those currently available in the art are not usually practicable.

Thus, there exists a need in the art for methods of efficiently altering predetermined endogenous genetic sequences by homologous pairing and homologous recombination in vivo by introducing one or more exogenous targeting polynucleotide(s) that efficiently and specifically homologously pair with a predetermined endogenous DNA sequence. There exists a need in the art for high-efficiency gene targeting, so as to avoid complex in vitro selection protocols (e.g., *neo* gene selection with G418) which are of limited utility for in vivo gene therapy on affected individuals

SUMMARY OF THE INVENTION

It is an object of the present invention to provide methods for targeting an exogenous polynucleotide to a predetermined endogenous DNA target sequence in a target cell with high efficiency and with sequence specificity. Exogenous polynucleotides, are localized (or targeted) to one or more predetermined DNA target sequence(s) by homologous pairing in

vivo. Such targeted homologous pairing of exogenous polynucleotides to endogenous DNA sequences in vivo may be used: (1) to target chemical substituents in a sequence-specific manner in vivo, (2) to correct or to generate genetic mutations in endogenous DNA sequences by homologous recombination and/or gene conversion, (3) to produce homologously targeted transgenic organisms, including animals and plants at high efficiency, and (4) in other applications (e.g., targeted drug delivery) based on in vivo homologous pairing. Some embodiments of the invention employ targeted exogenous polynucleotides to correct endogenous mutant gene alleles in human cells; the invention provides methods and compositions for correcting disease alleles involved in producing human genetic diseases, such as inherited genetic diseases (e.g., cystic fibrosis) and neoplasia (e.g., neoplasms induced by somatic mutation of an oncogene or tumor suppressor gene, such as p53, or viral genes associated with neoplasia, such as HBV genes).

In one embodiment, at least one exogenous polynucleotide is targeted to a predetermined endogenous DNA sequence and alters the endogenous DNA sequence, such as a chromosomal DNA sequence, typically by targeted homologous recombination within and/or flanking the predetermined endogenous DNA sequence. Generally, two complementary exogenous polynucleotides are used for targeting an endogenous DNA sequence. Typically, the targeting polynucleotide(s) are introduced simultaneously or contemporaneously with one or more recombinase species. Alternatively, one or more recombinase species may be produced in vivo by expression of a heterologous expression cassette in a cell containing the preselected target DNA sequence.

It is another object of the invention to provide methods whereby at least one exogenous polynucleotide containing a chemical substituent can be targeted to a predetermined endogenous DNA sequence in a metabolically-active or intact living target cell, permitting sequence-specific targeting of chemical substituents such as, for example cross-linking agents, metal chelates (e.g., iron/EDTA chelate for iron catalyzed cleavage), topoisomerases, endonucleases, exonucleases, ligases, phosphodiesterases, photodynamic porphyrins, free-radical generating drugs, chemotherapeutic drugs (e.g., adriamycin, doxirubicin), intercalating agents, base-modification agents, immunoglobulin chains, oligonucleotides, and

other substituents. The methods of the invention can be used to target such a chemical substituent to a predetermined DNA sequence by homologous pairing for various applications, for example: producing sequence-specific strand scission(s), producing sequence-specific chemical modifications (e.g., base methylation, strand cross-linking), producing sequence-specific localization of polypeptides (e.g., topoisomerases, helicases, proteases), producing sequence-specific localization of polynucleotides (e.g., loading sites for transcription factors and/or RNA polymerase), and other applications.

It is another object of the present invention to provide methods for correcting a genetic mutation in an endogenous DNA target sequence, such as a sequence encoding an RNA or a protein. For example, the invention can be used to correct genetic mutations, such as base substitutions, additions, and/or deletions, by converting a mutant DNA sequence that encodes a non-functional, dysfunctional, and/or truncated polypeptide into a corrected DNA sequence that encodes a functional polypeptide (e.g., has a biological activity such as an enzymatic activity, hormone function, or other biological property). The methods and compositions of the invention may also be used to correct genetic mutations or dysfunctional alleles with genetic lesions in non-coding sequences (e.g., promoters, enhancers, silencers, origins of replication, splicing signals). In contradistinction, the invention also can be used to target DNA sequences for inactivating gene expression; a targeting polynucleotide can be employed to make a targeted base substitution, addition, and/or deletion in a structural or regulatory endogenous DNA sequence to alter expression of one or more genes, typically by knocking out at least one allele of a gene (i.e., making a mutant, nonfunctional allele). The invention can also be used to correct disease alleles, such as a human or non-human animal CFTR gene allele associated with cystic fibrosis, by producing a targeted alteration in the disease allele to correct a disease-causing lesion (e.g., a deletion).

It is a further object of the invention to provide methods and compositions for high-efficiency gene targeting of human genetic disease alleles, such as a CFTR allele associated with cystic fibrosis or an LDL receptor allele associated with familial hypercholesterolemia. In one aspect of the invention, targeting polynucleotides having at least one associated recombinase are targeted to cells in vivo (i.e., in an intact animal) by exploiting the advantages of a

receptor-mediated uptake mechanism, such as an asialoglycoprotein receptor-mediated uptake process. In this variation, a targeting polynucleotide is associated with a recombinase and a cell-uptake component which enhances the uptake of the targeting polynucleotide-recombinase into cells of at least one cell type in an intact individual. For example, but not limitation, a cell-uptake component typically consists of: (l) a galactose-terminal (asialo-) glycoprotein (e.g., asialoorosomucoid) capable of being recognized and internalized by specialized receptors (asialoglycoprotein receptors) on hepatocytes in vivo, and (2) a polycation, such as poly-L-lysine, which binds to the targeting polynucleotide, usually by electrostatic interaction. Typically, the targeting polynucleotide is coated with recombinase and cell-uptake component simultaneously so that both recombinase and cell-uptake component bind to the targeting polynucleotide; alternatively, a targeting polynucleotide can be coated with recombinase prior to incubation with a cell-uptake component; alternatively the targeting polynucleotide can be coated with the cell-uptake component and introduced into cells contemporaneously with a separately delivered recombinase (e.g., by targeted liposomes containing one or more recombinase).

The invention also provides methods and compositions for diagnosis, treatment and prophylaxis of genetic diseases of animals, particularly mammals, wherein a recombinase and a targeting polynucleotide are used to produce a targeted sequence modification in a disease allele of an endogenous gene. The invention may also be used to produce targeted sequence modification(s) in a non-human animal, particularly a non-human mammal such as a mouse, which create(s) a disease allele in a non-human animal. Sequence-modified non-human animals harboring such a disease allele may provide useful models of human and veterinary disease(s). Alternatively, the methods and compositions of the invention can be used to provide nonhuman animals having homologously-targeted human disease alleles integrated into a non-human genome; such non-human animals may provide useful experimental models of human or other animal genetic disease, including neoplastic and other pathogenic diseases.

It is also an object of the invention to provide methods and compositions to introduce genes at a predetermined site in the chromosome. The invention may be used to introduce

heterologous cDNA and/or genomic DNA sequences into the chromosome of non-human animals to allow expression and/or production of the heterologous gene.

It is also an object of the invention to provide methods and compositions for recombinase-enhanced positioning of a targeting polynucleotide to a homologous sequence in an endogenous chromosome to form a stable multistrand complex, and thereby alter expression of a predetermined gene sequence by interfering with transcription of sequence(s) adjacent to the multistrand complex. Recombinase(s) are used to ensure correct homologous pairing and formation of a stable multistrand complex, which may include a double-D loop structure. For example, a targeting polynucleotide coated with a recombinase may homologously pair with an endogenous chromosomal sequence in a structural or regulatory sequence of a gene and form a stable multistrand complex which may: (1) constitute a significant physical or chemical obstacle to formation of or procession of an active transcriptional complex comprising at least an RNA polymerase, or (2) alter the local chromatin structure so as to alter the transcription rate of gene sequences within about 1 to 500 kilobases of the multistrand complex.

It is another object of the invention to provide methods and compositions for treating or preventing acquired human and animal diseases, particularly parasitic or viral diseases, such as human hepatitis B virus (HBV) hepatitis, by targeting viral gene sequences with a recombinase-associated targeting polynucleotide and thereby inactivating said viral gene sequences and inhibiting viral-induced pathology.

It is a further object of the invention to provide compositions that contain exogenous targeting polynucleotides, complementary pairs of targeting polynucleotides, chemical substituents of such polynucleotides, and recombinase proteins used in the methods of the invention. Such compositions may include a targeting or cell-uptake components to facilitate intracellular uptake of a targeting polynucleotide, especially for in vivo gene therapy and gene modification.

In accordance with the above objects, the present invention provides methods for targeting and altering, by homologous recombination, a pre-selected target nucleic acid sequence in a procaryotic cell to make a targeted sequence modification. The methods comprise introducing into at least one procaryotic cell at least one recombinase and at least two single-stranded targeting polynucleotides, each of which are substantially complementary to each other and comprise a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence.

In an additional aspect, the methods comprise adding to an extrachromosomal sequence at least one recombinase and at least two single-stranded targeting polynucleotides, each of which are substantially complementary to each other and comprise a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence to form an altered extrachromosomal sequence. The recombinase is then removed and the altered sequence is introduced into a target cell.

In a further embodiment, the present invention provides methods of generating a library of variant nucleic acid sequences of a pre-selected target nucleic acid sequence in an extrachromosomal sequence. The method comprises adding to an extrachromosomal sequence at least one recombinase and a plurality of pairs of single-stranded targeting polynucleotides, which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence. The plurality of pairs comprises a library of mismatches between the targeting polynucleotides and the target nucleic acid sequence, to form a library of altered extrachromosomal sequences.

In an additional aspect, the invention provides methods of generating a cellular library comprising variant nucleic acid sequences of a pre-selected target nucleic acid sequence. The methods comprise introducing into a population of target cells at least one recombinase and a plurality of pairs of single-stranded targeting polynucleotides which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid

sequence. The plurality of pairs comprises a library of mismatches between the targeting polynucleotides and the target nucleic acid sequence, to form said cellular library comprising variant nucleic acid sequences.

In a further aspect, the invention provides methods of generating a cellular library comprising variant nucleic acid sequences of a pre-selected target nucleic acid sequence in an extrachromosomal sequence of a target cell. The methods comprises adding to an extrachromosomal sequence at least one recombinase and a plurality of pairs of single-stranded targeting polynucleotides which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence. The plurality of pairs comprises a library of mismatches between the targeting polynucleotides and the target nucleic acid sequence, to form a plurality of altered extrachromosomal sequences. The recombinase is then removed and the altered sequences are introduced into a population of target cells to form the library of variant nucleic acid sequences.

The invention also provides compositions comprising at least one recombinase and a variant library comprising a plurality of pairs of single stranded targeting polynucleotides which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence. The plurality of pairs comprises a library of mismatches between the targeting polynucleotides and the target nucleic acid sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. Homologous targeting of recA-coated chromosome 1 alpha-satellite polynucleotides in metabolically active cell nuclei. The homologously targeted biotinylated polynucleotides were visualized by addition of FITC-avidin followed by washing to remove unbound FITC. Signals were visualized using a Zeiss Confocal Laser Scanning Microscope (CLSM-10) with 488 nm argon laser beam illumination for FITC-DNA detection. Top left - localized FITC-DNA signals in cell nucleus. Lower left - enhanced image of FITC-DNA signals in

cell nucleus. Upper right - image of FITC-DNA signals overlaid on phase image of nucleus. Lower right - phase image of center of cell nucleus showing nucleoli. Note: all images except lower right were photographed at same focus level (focus unchanged between these photos).

Figs. 2A, 2B, 2C, 2D, 2E, 2F, 2G, 2H, 2I, 2J, 2K, and 2L. RecA protein-mediated native FISH in metabolically active cell nuclei. Hep-2 cell nuclei from cells encapsulated in agarose were incubated with RecA-coated biotinylated p53 DNA (A-I) or RecA-coated biotinylated chromosome 1 satellite III DNA probes (K-L). Panels B-I show FISH signals in digital images from serial CLSM optical sections of FITC-labeled p53 probe DNA incubated in metabolically active Hep-2 nuclei. The phase image of a representative nucleous in shown in Panel A and was sectioned by CLSM. Digital images in Panels B-H were serially overlaid upon one another to produce the composite digital image shown in Panel I containing all three FITC labeled p53 FISH signals. The effect of cssDNA probe concentration and RecA protein on efficiency of native dsDNA hybridization in metabolically active nuclei is shown in Panel J. The percentage of labeled RecA coated or uncoated p53 cssDNA is shown as a function of the amount of p53 DNA probe per hybridization reaction. Closed circles show hybridization reactions with RecA-coated p53 cssDNA probe, open triangles show control reactions without RecA protein coating of p53 cssDNA probe. Panel K shows the FISH digital image in Panel L overlaid onto the phase image.

Fig. 3. Genetic map of mammalian expression lacZ plasmid pMC1lacXpA with an 11 base insertion in Xba linker site.

Fig. 4. Genetic map of mammalian expression lacZ plasmid pMC1lacpA, with insertion mutation.

Fig. 5. PCR products and primers from the lacZ (ß-galactosidase) gene sequence. The location of the 11 bp Xba linker is shown.

Fig. 6. Tests for alteration of an insertion mutation in the lacZ gene of a eukaryotic expression vector. NIH 3T3 cells were needle microinjected with five types of plasmids:

Two plasmids contained a wild-type ß-galactosidase gene (pMC1lacpa or pSV-ß-gal [Promega]); a plasmid with a mutant ß-gal gene (pMC1lacXpa); pMC1lacXpa plasmid reacted with an uncoated wild-type 276-mer DNA; or pMC1lacXpa plasmid reacted and D-looped with RecA-coated wild-type 276-mer DNA. The wild-type 276-mer DNA was either coated or not with RecA protein in a standard coating reaction protocol (Sena and Zarling, supra). Following a 10-min RecA coating reaction, the complementary RecA-coated single-stranded 276-mers were incubated at 37°C for 60 min. with the mutant target plasmid to allow hybrid formation. A 60 min incubation of the mutant target plasmid DNA with uncoated complementary single-stranded normal wild-type 276-mers was carried out as a control. The ß-galactosidase activity in needle microinjected cells using the wild-type plasmids is shown for comparison. On average, about 50% of the total microinjected cells survived. The numbers of surviving cells scoring blue with the mutant plasmid RecA-treated and non-RecA-treated samples (3, 4 and 5) were compared with fourfold χ^2 tests. The frequency of corrected blue cells in the RecA-treated sample (Sample 5; 6 out of 168) is significantly higher than that of either Sample 3 or Sample 4. The frequency of corrected RecA-treated blue cells in Sample 5 is significantly higher than that of Sample 4 at the 5% significance level ($\chi^2 = 3.76 > \chi^2_{0.05}$). The frequency of corrected RecA-treated blue cells in Sample 5 is significantly higher than that of Sample 3 at the 1% significance level ($\chi^2 = 6.28 > \chi^2_{0.01}$). When Samples 3 and 4 are combined and compared with Sample 5, the frequency of corrected blue cells in the RecA-treated Sample 5 is significantly higher than that of the combined sample at the 0.1% signficance level ($\chi^2 = 9.99 > \chi^2_{0.001}$).

Fig. 7A. Southern hybridization analysis of the 687-bp fragment amplified from genomic DNA. Electrophoretic migration of a 687-bp DNA fragment generated with primers CF1 and CF6 from genomic DNA of ∑CFTE29o-cells which were capillary needle-microinjected with the 491-nucleotide DNA fragment in the presence of recA (lane 2) or transfected as a protein-DNA-lipid complex where the 491-nucleotide fragments were coated with recA (+; lane 3). The control DNA was amplified from nontransfected ∑CFTE29o-cultures (lane 1).

Fig. 7B. Autoradiographic analysis of DNA transferred to Gene Screen Plus filters and hybridized with a ^{32}P-labeled oligonucleotide specific for normal exon 10 sequences in the

region of the ΔF508 mutation. Cells transfected by micro-injection or protein-lipid-DNA complexes both were positive for homologous targeting, whereas control cells were not.

Fig. 8A. Analysis of DNA from cells electroporated or transfected with DNA encapsulated in a protein-lipid complex. Allele-specific PCR amplification of the 687/684 bp DNA fragment amplified in the first round with primers CF1 and oligo N (N) or oligo ΔF (ΔF). Ethidium bromide-stained 300 bp DNA fragment separated by electrophoresis in a 1% agarose gel. The DNA in each lane is as follows: lane 1, 100-bp marker DNA; lane 2, control 16HBE14o-cell DNA amplified with the CF1/N primer pair; lane 3, nontransfected ΣCFTE29o-cell DNA amplified with CF1/N primers; lane 4, nontransfected ΣCFTE29o-cell DNA amplified with CF1/ΔF primers; lane 5, DNA from ΣCFTE29o-cells electroporated with recA-coated 491-nucleotide fragments and amplified with CF1/N primers; lane 6, DNA from ΣCFTE29o-cells transfected with recA-coated 491-nucleotide fragment encapsulated in a protein-lipid complex and amplified with CF1/N primers.

Fig. 8B. Autoradiographic analysis of the DNA in Fig. 11A transferred to Gene Screen Plus filters and hybridized with ^{32}P-labeled oligo N probe. Samples in lanes 1-5 for the autoradiographic analysis are equivalent to samples in lanes 2-6 in Fig. 11A.

Fig. 9. PCR analysis of ΣCFTE29o-genomic DNA reconstructed with the addition of 2 x 10^5 copies of recA-coated 491-nucleotide DNA fragments per microgram of genomic DNA. This number of DNA fragments represents the total number of DNA copies microinjected into cells and tests whether the 491-nucleotide fragment can act as a primer for the 687/684-bp fragment amplification. DNA was amplified as described in Fig. 8A. When the second round of amplification was conducted with CF1 and oligo N primers (lane 2), the 300-bp DNA band was not detected when aliquots of the amplification reaction were separated electrophoretically. Amplification of the ΣCFTE29o/491 bp DNA fragment with the CF1/oligo ΔF primer pair produced a 299-bp DNA product (lane 1). Marker DNA is in lane 3.

Figure 10 depicts the scheme for the recombination assay used in Example 4.

Fig. 11 shows RecA mediated cssDNA targeting to dsDNA with deletions produces a mixed population of probe:target hybrids. The biotinylated cssDNA probes were denatured and coated with RecA at 37°C as described in Material . The reaction mixture was incubated for 60 minutes at 37°C. All reactions were stopped by deproteinization with 1.2% SDS and separated by electrophoresis on a 20 cm X 25 cm 1% agarose gel. The gel was run overnight at 30V then blotted onto a positively charged TropilonPlus (TROPIX) membrane. The DNA was monitored for the presence of unhybridized probe or probe:target hybrids using an alkaline phosphatase based chemiluminescent detection of biotin. When the membranes were exposed to X-ray film and developed, it is evident that cssDNA probes will hybridize to dsDNA targets which are completely homologous as well as dsDNA targets which contain a deletion (lanes 3 and 6, respectively). RecA mediated cssDNA targeting to completely homologous dsDNA (pRD.0) forms a probe:target hybrid whose electrophoretic mobility is comparable to the electrophoretic mobility of completely relaxed Form I DNA which is similar to the mobility of Form II DNA (lanes 3, 8, and 10), referred to as the rI* hybrid. RecA hybridization of mediated cssDNA to dsDNA containing a 59 base pair deletion (pRD.59), a probe:target hybrid that migrates to a position similar to Form I DNA (lane 6), referred to as the I* hybrid.

Fig. 12 shows data for the enhanced homologous recombination (EHR) of cssDNA probe:target hybrids in E. coli, as per Example 4. The homologously targeted probe:target hybrids have enhanced homologous recombination frequencies in recombination proficient cells. cssDNA probe:target hybrids formed as in the legend of Figure 11 were introduced into RecA+ and RecA- E. coli as in Figure 12. The molar ratio of cssDNA probe:target in the in vitro targeting reaction varied from 1:1 to 1:5.6. The % recombinant/total colonies is the percentage blue colonies in the total population of ampicillin-resistant colonies. Groups with 0% recombinants did not produce any blue colonies in at least 10^5 plated colonies. Plasmid DNA was isolated from blue colonies that were serially propagated for three generations to determine if homologous recombination stably occurred in the lacZ gene.

Fig. 13 shows double D-loop hybrids with internal homology clamps. A) Duplex target DNA (thin line) is completely homologous to the cssDNA probe (thick) and each probe strand can pair with its complementary strand in the target. B) Duplex target has a deletion with respect to the cssDNA probe. The deleted region is indicated with a dashed line. The region of the cssDNA probes homologous to the deleted region in the target can re-pair with each other forming a stable hybrid complex. C) Duplex target has an insertion (dashed line) with respect to the cssDNA probe. Structures on the left show the re-annealing of cssDNA probe or target strands to form internal homology clamps. Structures on the right show the presence of unpaired regions in comparable single D-loop hybrids.

Figs. 14A and 14B. Figure 14A depicts the Maps of Plasmids pRD.0 and pRD.59. Relative positions of cssDNA probes IP290 and CP443, PCR primers 1A and 4B, restriction endonuclease sites EcoRI, ScaI, and DraI are indicated. The alpha peptide sequence of the LacZ gene is indicated. Note the deletion (Δ) in pRD.59 is approximately equidistant from the ends of primers 1A and 4B. Figure 14B). Time Course for cssDNA probe:target hybrid formation with linear dsDNA targets. Biotinylated, RecA coated cssDNA probe IP290 was hybridized as described to ScaI-digested plasmids pRD.0 and pRD.59 carrying 0 or 59 bp deletion, respectively at the EcoR1 site in pRD.0. Probe IP290 is completely homologous to pRD.0, but has a 59 bp insertion with respect to pRD.59.

Fig. 15 depicts the formation of cssDNA probe target hybrids in linear dsDNA targets containing small deletions. A) Plasmid constructs and probes used in this study. A series of plasmids with defined deletions were constructed from the EcoR1 site of pRD.0 (pbluescriptIISK+ (Stratagene) as described in Example 5. Each plasmid is named for the size of the deletion, as indicated on the left. A series of cssDNA probes were labelled and constructed by PCR from various primers which flank the deleted region. Probes were made from either pRD.0 or the deleted plasmids and named for the size of the probe when made from pRD.0 (2960 bp). For example, p527 is 527 bp long. When the cssDNA probes are produced from pRD.0 and targeted to plasmids containing deletions, the probe is called IP527 to indicate that the probe has an insertion with respect to the target. When the probe is made from one of the targets with a deletion and then, targeted to pRD.0, the probe is called DP527

to indicate that the probe has a deletion with respect to pRD.0. Control probe CP443 is made from a region of pRD.0 that does not contain any insertions or deletions. The limits of the deleted regions in the plasmid DNA target are indicated by dashed line and the size limits of cssDNA probes are indicated by solid lines. B) Biotinylated cssDNA probes IP527, IP407, and CP443 were coated with RecA protein and hybridized at 37°C to a series of linear duplex DNA targets containing deletions ranging in size from 0 to 447 bp. The products of the targeting reaction were deproteinized and separated on a 1 % TAE-agarose gel and then transferred to nylon membranes as described in Example 5. Biotinylated DNA was detected with a chemiluminescent substrate as described. The extent of hybrid product formation of FormIII DNA targets was determined by densitometry of the autoradiographs. The relative amount of hybrid formed between RecA coated cssDNA probes IP527 and IP407 is shown in (B). Error bars are indicated. The amount of probe:target hybrids formed with each target DNA was normalized by the amount of probe:target hybrids formed with control probe CP443 which hybridizes to the target away from the deletion site. Examples of the cssDNA probe:target hybrid formed with linear targets is shown in the autoradiogram (C). In Fig. 15(D) the difference in the percent hybrid formation between cssDNA probes IP527 and IP407 are plotted from the data shown in (B).

Fig. 16 depicts that insertions and deletions have the same effect on the relative efficiency of probe:target hybrid formation. RecA-coated cssDNA probes IP215 made from pRD.0 was targeted to Sca1-digests of plasmids pRD.0, pRD.8, pRD.25, and pRD.59 and compared to similar reactions of DP215 cssDNA probes made from pRD.0, pRD.8, pRD.25, and pRD.59 and targeted to pRD.0. The effect of insertions in the cssDNA probe (dark line) is compared with deletions in the cssDNA probe (shaded line) of the same size. The relative level of hybrid formation for each cssDNA probe with a heterologous target is normalized by the level of hybridization with the homologous target, respectively. The data represents an average of three experiments. Error bars are indicated.

Figs 17A, 17B and 17C. Figure 17A depict the formation of stable Double-D-Loop hybrids in linear dsDNA targets containing large deletions. Biotinylated cssDNA probe IP1246 was coated with RecA protein and targeted to Sca1 digests of the indicated plasmids as described

herein. The relative amount of hybrid formation formed between RecA-coated cssDNA probes and plasmids with deletions ranging from 0-967 bp was normalized to the amount of probe:target hybrids formed with control probe CP443. Autoradiograph (17A) shows the biotinylated cssDNA probes or probe:target hybrids. The position of the untargeted ScaI-digested (FormIII) marker for each of the plasmids are indicated on the right. The relative level of hybrid formation (B) of each of the bands in (A) was normalized to the level of hybrid formation with control cssDNA probe CP443 as described herein. The relative position of the cssDNA probes with respect to the position of the deletion in the target DNA is shown in (C).

Figs. 18A, 18B, 18C and 18D depict the formation of restriction endonuclease sites in probe:target hybrids. The probe:target hybrids formed between probe IP290 and pRD.0 and pRD.59 targets were deproteinized by extraction with chloroform:phenol:isoamyl alcohol and chloroform. Restriction enzyme treated DNA samples were incubated with EcoRI for three hours before separation on a 1% agarose gel and transferred onto a nylon membrane. The ethidium bromide stained DNA of the products of the targeting reactions formed between cssDNA probe IP290 and circular plasmid targets pRD.0 or pRD.59 (A and B) and autoradiographs showing the positions of biotinylated cssDNA probe:target hybrids (C and D) are shown. The positions of form I and form III markers of pRD.0 are shown on the right. The positions of the pRD59 hybrids I* (form I) and rI* (relaxed) are shown on the left.

Fig. 19 depicts the thermal stability of relaxed and non-relaxed probe:target hybrids. The RecA mediated cssDNA targeting reaction was performed with the cssDNA probe IP290 and the dsDNA target pRD.59 as described herein. The probe:target hybrids were deproteinized with 1.2% SDS and then incubated for 5 minutes at the indicated temperatures. The thermally melted products were then separated on a 1% agarose gel and blotted onto a positively charged Tropilon membrane. Autoradiograph shows the position of biotinylated cssDNA probe:target hybrids I* (formI) and rI* (relaxed) as shown on the left.

Figs. 20A and 20B. The organization of the mouse OTC gene. Sequence of cssDNA probes and PCR primers used in this study are indicated. Sizes of the exons in basepairs are

indicated. The relative position of PCR primers M9, M8 and M11 are shown. B) Map of plasmid pTAOTC1. A 250 bp fragment containing the normal OTC exon4 sequence and surrounding introns were cloned into the EcoRV site of pbluescript SK (+) (Stratagene).

Fig. 21. Sequence analysis of exon4 of the mouse OTC gene in founder mice. PCR amplification of genomic DNA from tail biopsies of a pool of all of the homozygous (spf-ash/spf-ash) females used as egg donors and each indicated individual founder mice were sequenced using cycle sequencing with the M11 primer (Cyclist kit, Stratagene). The DNA sequence surrounding the spf-ash locus (arrow) in the OTC gene is shown.

Fig. 22. Germline transmission of OTC+ allele corrected by EHR. The inheritance patterns of the OTC alleles are depicted. Legend indicates the genotype and/or phenotype of the F0, F1, and F2 mice produced from microinjected zygotes obtained from the cross of homozygous (spf-ash/spf-ash) mutant females and normal males (top). The genotype of F0 and F1 animals were determined by DNA sequencing and the typing of F2 animals as deduced by phenotype. Control cross A of (hemizygous spf-ash/Y) mutant F0 male with normal (+/+) females and control cross B of heterozygous (spf-ash/+) F1 females with a normal male are indicated. The number below the boxes or circles indicate the total number of mice of each type produced from each cross. Total numbers of mice counted are representative of 2-4 litters. Mouse #213 and #1014 (noted by arrow) are F1 animals that carry a germline transmitted gene corrected allele from mosaic HR gene corrected male mouse #16.

Fig. 23. Germline transmission of corrected allele of F0 male #16. Pictures of F1 progeny from the cross of mouse #16 with homozygous (spf-ash/spf-ash) females (top). This cross produced several pups with spf-ash mutant phenotypes (middle) and one F1 pup (#1014) with a normal phenotype. Three views of mouse #1014 are shown (bottom). All of the F1 animals were two weeks old at the time of photography.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage (Immunology - A Synthesis, 2nd Edition, E.S. Golub and D.R. Green, Eds., Sinauer Associates, Sunderland, Massachusetts (1991), which is incorporated herein by reference).

By "nucleic acid", "oligonucleotide", and "polynucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487 (1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate, phosphorodithioate, O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm. J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are incorporated by reference). These modifications of the ribose-phosphate backbone or bases may be done to facilitate the addition of other moieties such as chemical constituents, including 2' O-methyl and 5' modified substituents, as discussed below, or to increase the stability and half-life of such molecules in physiological environments.

The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo-and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine and hypoxathanine, etc. Thus, for example, chimeric DNA-RNA molecules may be used such as described in Cole-Strauss et al., Science 273:1386 (1996) and Yoon et al., PNAS USA 93:2071 (1996), both of which are hereby incorporated by reference.

In general, the targeting polynucleotides may comprise any number of structures, as long as the changes do not substantially effect the functional ability of the targeting polynucleotide to result in homologous recombination. For example, recombinase coating of alternate structures should still be able to occur.

As used herein, the terms "predetermined endogenous DNA sequence" and "predetermined target sequence" refer to polynucleotide sequences contained in a target cell. Such sequences include, for example, chromosomal sequences (e.g., structural genes, regulatory sequences including promoters and enhancers, recombinatorial hotspots, repeat sequences, integrated proviral sequences, hairpins, palindromes), episomal or extrachromosomal sequences (e.g., replicable plasmids or viral replication intermediates) including chloroplast and mitochondrial DNA sequences. By "predetermined" or "pre-selected" it is meant that the target sequence may be selected at the discretion of the practitioner on the basis of known or predicted sequence information, and is not constrained to specific sites recognized by certain site-specific recombinases (e.g., FLP recombinase or CRE recombinase). In some embodiments, the predetermined endogenous DNA target sequence will be other than a naturally occurring germline DNA sequence (e.g., a transgene, parasitic, mycoplasmal or viral sequence). An exogenous polynucleotide is a polynucleotide which is transferred into a target cell but which has not been replicated in that host cell; for example, a virus genome polynucleotide that enters a cell by fusion of a virion to the cell is an exogenous polynucleotide, however, replicated copies of the viral polynucleotide subsequently made in the infected cell are endogenous sequences (and may, for example, become integrated into a

cell chromosome). Similarly, transgenes which are microinjected or transfected into a cell are exogenous polynucleotides, however integrated and replicated copies of the transgene(s) are endogenous sequences.

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., may be similar or identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. As outlined below, preferably, the homology is at least 70%, preferably 85%, and more preferably 95% identical. Thus, the complementarity between two single-stranded targeting polynucleotides need not be perfect. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is perfectly complementary to a reference sequence "GTATA".

The terms "substantially corresponds to" or "substantial identity" or "homologous" as used herein denotes a characteristic of a nucleic acid sequence, wherein a nucleic acid sequence has at least about 70 percent sequence identity as compared to a reference sequence, typically at least about 85 percent sequence identity, and preferably at least about 95 percent sequence identity as compared to a reference sequence. The percentage of sequence identity is calculated excluding small deletions or additions which total less than 25 percent of the reference sequence. The reference sequence may be a subset of a larger sequence, such as a portion of a gene or flanking sequence, or a repetitive portion of a chromosome. However, the reference sequence is at least 18 nucleotides long, typically at least about 30 nucleotides long, and preferably at least about 50 to 100 nucleotides long. "Substantially complementary" as used herein refers to a sequence that is complementary to a sequence that substantially corresponds to a reference sequence. In general, targeting efficiency increases with the length of the targeting polynucleotide portion that is substantially complementary to a reference sequence present in the target DNA.

"Specific hybridization" is defined herein as the formation of hybrids between a targeting polynucleotide (e.g., a polynucleotide of the invention which may include substitutions, deletion, and/or additions as compared to the predetermined target DNA sequence) and a predetermined target DNA, wherein the targeting polynucleotide preferentially hybridizes to the predetermined target DNA such that, for example, at least one discrete band can be identified on a Southern blot of DNA prepared from target cells that contain the target DNA sequence, and/or a targeting polynucleotide in an intact nucleus localizes to a discrete chromosomal location characteristic of a unique or repetitive sequence. In some instances, a target sequence may be present in more than one target polynucleotide species (e.g., a particular target sequence may occur in multiple members of a gene family or in a known repetitive sequence). It is evident that optimal hybridization conditions will vary depending upon the sequence composition and length(s) of the targeting polynucleotide(s) and target(s), and the experimental method selected by the practitioner. Various guidelines may be used to select appropriate hybridization conditions (see, Maniatis et al., Molecular Cloning: A Laboratory Manual (1989), 2nd Ed., Cold Spring Harbor, N.Y. and Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA.; which are incorporated herein by reference. Methods for hybridizing a targeting polynucleotide to a discrete chromosomal location in intact nuclei are provided herein in the Detailed Description.

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring.

A metabolically-active cell is a cell, comprising an intact nucleoid or nucleus, which, when provided nutrients and incubated in an appropriate medium carries out DNA synthesis and RNA for extended periods (e.g., at least 12-24 hours). Such metabolically-active cells are typically undifferentiated or differentiated cells capable or incapable of further cell division (although non-dividing cells many undergo nuclear division and chromosomal replication), although stem cells and progenitor cells are also metabolically-active cells.

As used herein, the term "disease allele" refers to an allele of a gene which is capable of producing a recognizable disease. A disease allele may be dominant or recessive and may produce disease directly or when present in combination with a specific genetic background or pre-existing pathological condition. A disease allele may be present in the gene pool or may be generated de novo in an individual by somatic mutation. For example and not limitation, disease to alleles include: activated oncogenes, a sickle cell anemia allele, a Tay-Sachs allele, a cystic fibrosis allele, a Lesch-Nyhan allele, a retinoblastoma-susceptibility allele, a Fabry's disease allele, and a Huntington's chorea allele. As used herein, a disease allele encompasses both alleles associated with human diseases and alleles associated with recognized veterinary diseases. For example, the ΔF508 CFTR allele is a human disease allele which is associated with cystic fibrosis in North Americans.

As used herein, the term "cell-uptake component" refers to an agent which, when bound, either directly or indirectly, to a targeting polynucleotide, enhances the intracellular uptake of the targeting polynucleotide into at least one cell type (e.g., hepatocytes). A cell-uptake component may include, but is not limited to, the following: specific cell surface receptors such as a galactose-terminal (asialo-) glycoprotein capable of being internalized into hepatocytes via a hepatocyte asialoglycoprotein receptor, a polycation (e.g., poly-L-lysine), and/or a protein-lipid complex formed with the targeting polynucleotide. Various combinations of the above, as well as alternative cell-uptake components will be apparent to those of skill in the art and are provided in the published literature.

DETAILED DESCRIPTION

Generally, the nomenclature used hereafter and the laboratory procedures in cell culture, molecular genetics, and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for recombinant nucleic acid methods, polynucleotide synthesis, cell culture, and transgenesis. Generally enzymatic reactions, oligonucleotide synthesis, oligonucleotide modification, and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are generally performed according to conventional methods in the

art and various general references which are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

Transgenic mice are derived according to Hogan, et al., "Manipulating the Mouse Embryo: A Laboratory Manual", Cold Spring Harbor Laboratory (1988) which is incorporated herein by reference.

Embryonic stem cells are manipulated according to published procedures (Teratocarcinomas and embryonic stem cells: a practical approach, E.J. Robertson, ed., IRL Press, Washington, D.C., 1987; Zjilstra et al., Nature 342:435-438 (1989); and Schwartzberg et al., Science 246:799-803 (1989), each of which is incorporated herein by reference).

Zygotes are manipulated according to known procedures; for example see U.S. Patent No. 4,873,191, Brinster et al., PNAS 86:7007 (1989); Susulic et al., J. Biol. Chem. 49:29483 (1995), and Cavard et al., Nucleic Acids Res. 16:2099 (1988), hereby incorporated by reference.

Oligonucleotides can be synthesized on an Applied Bio Systems oligonucleotide synthesizer according to specifications provided by the manufacturer. Modified oligonucleotides and peptide nucleic acids are made as is generally known in the art.

The present invention provides methods for targeting and altering, by homologous recombination, a pre-selected target nucleic acid sequence in a target cell, to make targeted sequence modifications. The methods comprise introducing into the target cells a recombinase and at least two single-stranded targeting polynucleotides which are substantially complementary to each other. The targeting polynucleotides each comprise at least one homology clamp that substantially corresponds to or is substantially complementary to the preselected target nucleic acid sequence. The target cells are then screened to identify target cells containing the targeted sequence modification.

Targeting Polynucleotides

Targeting polynucleotides may be produced by chemical synthesis of oligonucleotides, nick-translation of a double-stranded DNA template, polymerase chain-reaction amplification of a sequence (or ligase chain reaction amplification), purification of prokaryotic or target cloning vectors harboring a sequence of interest (e.g., a cloned cDNA or genomic clone, or portion thereof) such as plasmids, phagemids, YACs, cosmids, bacteriophage DNA, other viral DNA or replication intermediates, or purified restriction fragments thereof, as well as other sources of single and double-stranded polynucleotides having a desired nucleotide sequence. Targeting polynucleotides are generally ssDNA or dsDNA, most preferably two complementary single-stranded DNAs.

Targeting polynucleotides are generally at least about 2 to 100 nucleotides long, preferably at least about 5- to 100 nucleotides long, at least about 250 to 500 nucleotides long, more preferably at least about 500 to 2000 nucleotides long, or longer; however, as the length of a targeting polynucleotide increases beyond about 20,000 to 50,000 to 400,000 nucleotides, the efficiency or transferring an intact targeting polynucleotide into the cell decreases. The length of homology may be selected at the discretion of the practitioner on the basis of the sequence composition and complexity of the predetermined endogenous target DNA sequence(s) and guidance provided in the art, which generally indicates that 1.3 to 6.8 kilobase segments of homology are preferred (Hasty et al. (1991) Molec. Cell. Biol. 11: 5586; Shulman et al. (1990) Molec. Cell. Biol. 10: 4466, which are incorporated herein by reference). Targeting polynucleotides have at least one sequence that substantially corresponds to, or is substantially complementary to, a predetermined endogenous DNA sequence (i.e., a DNA sequence of a polynucleotide located in a target cell, such as a chromosomal, mitochondrial, chloroplast, viral, episomal, or mycoplasmal polynucleotide). Such targeting polynucleotide sequences serve as templates for homologous pairing with the predetermined endogenous sequence(s), and are also referred to herein as homology clamps. In targeting polynucleotides, such homology clamps are typically located at or near the 5' or 3' end, preferably homology clamps are internally or located at each end of the polynucleotide (Berinstein et al. (1992) Molec. Cell. Biol. 12: 360, which is incorporated herein by

reference). Without wishing to be bound by any particular theory, it is believed that the addition of recombinases permits efficient gene targeting with targeting polynucleotides having short (i.e., about 50 to 1000 basepair long) segments of homology, as well as with targeting polynucleotides having longer segments of homology.

Therefore, it is preferred that targeting polynucleotides of the invention have homology clamps that are highly homologous to the predetermined target endogenous DNA sequence(s), most preferably isogenic. Typically, targeting polynucleotides of the invention have at least one homology clamp that is at least about 18 to 35 nucleotides long, and it is preferable that homology clamps are at least about 20 to 100 nucleotides long, and more preferably at least about 100-500 nucleotides long, although the degree of sequence homology between the homology clamp and the targeted sequence and the base composition of the targeted sequence will determine the optimal and minimal clamp lengths (e.g., G-C rich sequences are typically more thermodynamically stable and will generally require shorter clamp length). Therefore, both homology clamp length and the degree of sequence homology can only be determined with reference to a particular predetermined sequence, but homology clamps generally must be at least about 12 nucleotides long and must also substantially correspond or be substantially complementary to a predetermined target sequence. Preferably, a homology clamp is at least about 12, and preferably at least about 50 nucleotides long and is identical to or complementary to a predetermined target sequence. Without wishing to be bound by a particular theory, it is believed that the addition of recombinases to a targeting polynucleotide enhances the efficiency of homologous recombination between homologous, nonisogenic sequences (e.g., between an exon 2 sequence of a albumin gene of a Balb/c mouse and a homologous albumin gene exon 2 sequence of a C57/BL6 mouse), as well as between isogenic sequences.

The formation of heteroduplex joints is not a stringent process; genetic evidence supports the view that the classical phenomena of meiotic gene conversion and aberrant meiotic segregation result in part from the inclusion of mismatched base pairs in heteroduplex joints, and the subsequent correction of some of these mismatched base pairs before replication. Observations on recA protein have provided information on parameters that affect the

discrimination of relatedness from perfect or near-perfect homology and that affect the inclusion of mismatched base pairs in heteroduplex joints. The ability of recA protein to drive strand exchange past all single base-pair mismatches and to form extensively mismatched joints in superhelical DNA reflect its role in recombination and gene conversion. This error-prone process may also be related to its role in mutagenesis. RecA-mediated pairing reactions involving DNA of ϕX174 and G4, which are about 70 percent homologous, have yielded homologous recombinants (Cunningham et al. (1981) Cell 24: 213), although recA preferentially forms homologous joints between highly homologous sequences, and is implicated as mediating a homology search process between an invading DNA strand and a recipient DNA strand, producing relatively stable heteroduplexes at regions of high homology. Accordingly, it is the fact that recombinases can drive the homologous recombination reaction between strands which are significantly, but not perfectly, homologous, which allows gene conversion and the modification of target sequences. Thus, targeting polynucleotides may be used to introduce nucleotide substitutions, insertions and deletions into an endogeneous DNA sequence, and thus the corresponding amino acid substitutions, insertions and deletions in proteins expressed from the endogeneous DNA sequence.

In a preferred embodiment, two substantially complementary targeting polynucleotides are used. In one embodiment, the targeting polynucleotides form a double stranded hybrid, which may be coated with recombinase, although when the recombinase is recA, the loading conditions may be somewhat different from those used for single stranded nucleic acids.

In a prefered embodiment, two substantially complementary single-stranded targeting polynucleotides are used. The two complementary single-stranded targeting polynucleotides are usually of equal length, although this is not required. However, as noted below, the stability of the four strand hybrids of the invention is putatively related, in part, to the lack of significant unhybridized single-stranded nucleic acid, and thus significant unpaired sequences are not preferred. Furthermore, as noted above, the complementarity between the two targeting polynucleotides need not be perfect. The two complementary single-stranded targeting polynucleotides are simultaneously or contemporaneously introduced into a target

cell harboring a predetermined endogenous target sequence, generally with at lease one recombinase protein (e.g., recA). Under most circumstances, it is preferred that the targeting polynucleotides are incubated with recA or other recombinase prior to introduction into a target cell, so that the recombinase protein(s) may be "loaded" onto the targeting polynucleotide(s), to coat the nucleic acid, as is described below. Incubation conditions for such recombinase loading are described infra, and also in U.S.S.N. 07/755,462, filed 4 September 1991; U.S.S.N. 07/910,791, filed 9 July 1992; and U.S.S.N. 07/520,321, filed 7 May 1990, each of which is incorporated herein by reference. A targeting polynucleotide may contain a sequence that enhances the loading process of a recombinase, for example a recA loading sequence is the recombinogenic nucleation sequence poly[d(A-C)], and its complement, poly[d(G-T)]. The duplex sequence poly[d(A-C)•d(G-T)$_n$, where n is from 5 to 25, is a middle repetitive element in target DNA.

There appears to be a fundamental difference in the stability of RecA-protein-mediated D-loops formed between one single-stranded DNA (ssDNA) probe hybridized to negatively supercoiled DNA targets in comparison to relaxed or linear duplex DNA targets. Internally located dsDNA target sequences on relaxed linear DNA targets hybridized by ssDNA probes produce single D-loops, which are unstable after removal of RecA protein (Adzuma, Genes Devel. 6:1679 (1992); Hsieh et al, PNAS USA 89:6492 (1992); Chiu et al., Biochemistry 32:13146 (1993)). This probe DNA instability of hybrids formed with linear duplex DNA targets is most probably due to the incoming ssDNA probe W-C base pairing with the complementary DNA strand of the duplex target and disrupting the base pairing in the other DNA strand. The required high free-energy of maintaining a disrupted DNA strand in an unpaired ssDNA conformation in a protein-free single-D-loop apparently can only be compensated for either by the stored free energy inherent in negatively supercoiled DNA targets or by base pairing initiated at the distal ends of the joint DNA molecule, allowing the exchanged strands to freely intertwine.

However, the addition of a second complementary ssDNA to the three-strand-containing single-D-loop stabilizes the deproteinized hybrid joint molecules by allowing W-C base pairing of the probe with the displaced target DNA strand. The addition of a second

RecA-coated complementary ssDNA (cssDNA) strand to the three-strand containing single D-loop stabilizes deproteinized hybrid joints located away from the free ends of the duplex target DNA (Sena & Zarling, Nature Genetics 3:365 (1993); Revet et al. J. Mol. Biol. 232:779 (1993); Jayasena and Johnston, J. Mol. Bio. 230:1015 (1993)). The resulting four-stranded structure, named a double D-loop by analogy with the three-stranded single D-loop hybrid has been shown to be stable in the absence of RecA protein. This stability likely occurs because the restoration of W-C basepairing in the parental duplex would require disruption of two W-C basepairs in the double-D-loop (one W-C pair in each heteroduplex D-loop). Since each base-pairing in the reverse transition (double-D-loop to duplex) is less favorable by the energy of one W-C basepair, the pair of cssDNA probes are thus kinetically trapped in duplex DNA targets in stable hybrid structures. The stability of the double-D loop joint molecule within internally located probe:target hybrids is an intermediate stage prior to the progression of the homologous recombination reaction to the strand exchange phase. The double D-loop permits isolation of stable multistranded DNA recombination intermediates.

In addition, when the targeting polynucleotides are used to generate insertions or deletions in an endogeneous nucleic acid sequence, the use of two complementary single-stranded targeting polynucleotides allows the use of internal homology clamps as depicted in Figure 13. The use of internal homology clamps allows the formation of stable deproteinized cssDNA:probe target hybrids with homologous DNA sequences containing either relatively small or large insertions and deletions within a homologous DNA target. Without being bound by theory, it appears that these probe:target hybrids, with heterologous inserts in the cssDNA probe, are stabilized by the re-annealing of cssDNA probes to each other within the double-D-loop hybrid, forming a novel DNA structure with an internal homology clamp. Similarly stable double-D-loop hybrids formed at internal sites with heterologous inserts in the linear DNA targets (with respect to the cssDNA probe) are equally stable. Because cssDNA probes are kinetically trapped within the duplex target, the multi-stranded DNA intermediates of homologous DNA pairing are stabilized and strand exchange is facilitated.

In a preferred embodiment, the length of the internal homology clamp (i.e. the length of the insertion or deletion) is from about 1 to 50% of the total length of the targeting

polynucleotide, with from about 1 to about 20% being preferred and from about 1 to about 10% being especially preferred, although in some cases the length of the deletion or insertion may be significantly larger. As for the targeting homology clamps, the complementarity within the internal homology clamp need not be perfect.

The invention may also be practiced with individual targeting polynucleotides which do not comprise part of a complementary pair. In each case, a targeting polynucleotide is introduced into a target cell simultaneously or contemporaneously with a recombinase protein, typically in the form of a recombinase coated targeting polynucleotide as outlined herein (i.e., a polynucleotide pre-incubated with recombinase wherein the recombinase is noncovalently bound to the polynucleotide; generally referred to in the art as a nucleoprotein filament).

A targeting polynucleotide used in a method of the invention typically is a single-stranded nucleic acid, usually a DNA strand, or derived by denaturation of a duplex DNA, which is complementary to one (or both) strand(s) of the target duplex nucleic acid. Thus, one of the complementary single stranded targeting polynucleotides is complementary to one strand of the endogeneous target sequence (i.e. Watson) and the other complementary single stranded targeting polynucleotide is complementary to the other strand of the endogeneous target sequence (i.e. Crick). The homology clamp sequence preferably contains at least 90-95% sequence homology with the target sequence, to insure sequence-specific targeting of the targeting polynucleotide to the endogenous DNA target. Each single-stranded targeting polynucleotide is typically about 50-600 bases long, although a shorter or longer polynucleotide may also be employed. Alternatively, targeting polynucleotides may be prepared in single-stranded form by oligonucleotide synthesis methods, which may first require, especially with larger targeting polynucleotides, formation of subfragments of the targeting polynucleotide, typically followed by splicing of the subfragments together, typically by enzymatic ligation.

Recombinase Proteins

Recombinases are proteins that, when included with an exogenous targeting polynucleotide, provide a measurable increase in the recombination frequency and/or localization frequency between the targeting polynucleotide and an endogenous predetermined DNA sequence. Thus, in a preferred embodiment, increases in recombination frequency from the normal range of 10^{-8} to 10^{-4}, to 10^{-4} to 10^{1}, preferably 10^{-3} to 10^{1}, and most preferably 10^{-2} to 10^{1}, may be acheived.

In the present invention, recombinase refers to a family of RecA-like recombination proteins all having essentially all or most of the same functions, particularly: (i) the recombinase protein's ability to properly bind to and position targeting polynucleotides on their homologous targets and (ii) the ability of recombinase protein/targeting polynucleotide complexes to efficiently find and bind to complementary endogenous sequences. The best characterized recA protein is from *E. coli*, in addition to the wild-type protein a number of mutant recA-like proteins have been identified (e.g., recA803; see Madiraju et al., PNAS USA 85(18):6592 (1988); Madiraju et al, Biochem. 31:10529 (1992); Lavery et al., J. Biol. Chem. 267:20648 (1992)). Further, many organisms have recA-like recombinases with strand-transfer activities (e.g., Fugisawa et al., (1985) Nucl. Acids Res. 13: 7473; Hsieh et al., (1986) Cell 44: 885; Hsieh et al., (1989) J. Biol. Chem. 264: 5089; Fishel et al., (1988) Proc. Natl. Acad. Sci. (USA) 85: 3683; Cassuto et al., (1987) Mol. Gen. Genet. 208: 10; Ganea et al., (1987) Mol. Cell Biol. 7: 3124; Moore et al., (1990) J. Biol. Chem. 19: 11108; Keene et al., (1984) Nucl. Acids Res. 12: 3057; Kimeic, (1984) Cold Spring Harbor Symp. 48: 675; Kmeic, (1986) Cell 44: 545; Kolodner et al., (1987) Proc. Natl. Acad. Sci. USA 84: 5560; Sugino et al., (1985) Proc. Natl. Acad. Sci. USA 85: 3683; Halbrook et al., (1989) J. Biol. Chem. 264: 21403; Eisen et al., (1988) Proc. Natl. Acad. Sci. USA 85: 7481; McCarthy et al., (1988) Proc. Natl. Acad. Sci. USA 85: 5854; Lowenhaupt et al., (1989) J. Biol. Chem. 264: 20568, which are incorporated herein by reference. Examples of such recombinase proteins include, for example but not limitation: recA, recA803, uvsX, and other recA mutants and recA-like recombinases (Roca, A. I. (1990) Crit. Rev. Biochem. Molec. Biol. 25: 415), sep1 (Kolodner et al. (1987) Proc. Natl. Acad. Sci. (U.S.A.) 84:5560; Tishkoff et al. Molec. Cell. Biol. 11:2593), RuvC (Dunderdale et al. (1991) Nature 354: 506), DST2, KEM1, XRN1 (Dykstra et al. (1991) Molec. Cell. Biol. 11:2583), STPα/DST1 (Clark et al. (1991) Molec.

Cell. Biol. 11:2576), HPP-1 (Moore et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88:9067), other target recombinases (Bishop et al. (1992) Cell 69: 439; Shinohara et al. (1992) Cell 69: 457); incorporated herein by reference. RecA may be purified from *E. coli* strains, such as *E. coli* strains JC12772 and JC15369 (available from A.J. Clark and M. Madiraju, University of California-Berkeley, or purchased commercially). These strains contain the recA coding sequences on a "runaway" replicating plasmid vector present at a high copy numbers per cell. The recA803 protein is a high-activity mutant of wild-type recA. The art teaches several examples of recombinase proteins, for example, from Drosophila, yeast, plant, human, and non-human mammalian cells, including proteins with biological properties similar to recA (i.e., recA-like recombinases), such as Rad51 from mammals and yeast, and Pk-rec (see Rashid et al., Nucleic Acid Res. 25(4):719 (1997), hereby incorporated by reference). In addition, the recombinase may actually be a complex of proteins, i.e. a "recombinosome". In addition, included within the definition of a recombinase are portions or fragments of recombinases which retain recombinase biological activity, as well as variants or mutants of wild-type recombinases which retain biological activity, such as the E. coli recA803 mutant with enhanced recombinase activity.

In a preferred embodiment, recA or rad51 is used. For example, recA protein is typically obtained from bacterial strains that overproduce the protein: wild-type *E. coli* recA protein and mutant recA803 protein may be purified from such strains. Alternatively, recA protein can also be purchased from, for example, Pharmacia (Piscataway, NJ).

RecA proteins, and its homologs, form a nucleoprotein filament when it coats a single-stranded DNA. In this nucleoprotein filament, one monomer of recA protein is bound to about 3 nucleotides. This property of recA to coat single-stranded DNA is essentially sequence independent, although particular sequences favor initial loading of recA onto a polynucleotide (e.g., nucleation sequences). The nucleoprotein filament(s) can be formed on essentially any DNA molecule and can be formed in cells (e.g., mammalian cells), forming complexes with both single-stranded and double-stranded DNA, although the loading conditions for dsDNA are somewhat different than for ssDNA.

Recombinase Coating of Targeting Polynucleotides

The conditions used to coat targeting polynucleotides with recombinases such as recA protein and ATPγS have been described in commonly assigned U.S.S.N. 07/910,791, filed 9 July 1992; U.S.S.N. 07/755,462, filed 4 September 1991; and U.S.S.N. 07/520,321, filed 7 May 1990, each incorporated herein by reference. The procedures below are directed to the use of E. coli recA, although as will be appreciated by those in the art, other recombinases may be used as well. Targeting polynucleotides can be coated using GTPγS, mixes of ATPγS with rATP, rGTP and/or dATP, or dATP or rATP alone in the presence of an rATP generating system (Boehringer Mannheim). Various mixtures of GTPγS, ATPγS, ATP, ADP, dATP and/or rATP or other nucleosides may be used, particularly preferred are mixes of ATPγS and ATP or ATPγS and ADP.

RecA protein coating of targeting polynucleotides is typically carried out as described in U.S.S.N. 07/910,791, filed 9 July 1992 and U.S.S.N. 07/755,462, filed 4 September 1991, which are incorporated herein by reference. Briefly, the targeting polynucleotide, whether double-stranded or single-stranded, is denatured by heating in an aqueous solution at 95-100°C for five minutes, then placed in an ice bath for 20 seconds to about one minute followed by centrifugation at 0°C for approximately 20 sec, before use. When denatured targeting polynucleotides are not placed in a freezer at -20°C they are usually immediately added to standard recA coating reaction buffer containing ATPγS, at room temperature, and to this is added the recA protein. Alternatively, recA protein may be included with the buffer components and ATPγS before the polynucleotides are added.

RecA coating of targeting polynucleotide(s) is initiated by incubating polynucleotide-recA mixtures at 37°C for 10-15 min. RecA protein concentration tested during reaction with polynucleotide varies depending upon polynucleotide size and the amount of added polynucleotide, and the ratio of recA molecule:nucleotide preferably ranges between about 3:1 and 1:3. When single-stranded polynucleotides are recA coated independently of their homologous polynucleotide strands, the mM and μM concentrations of ATPγS and recA, respectively, can be reduced to one-half those used with double-stranded targeting

polynucleotides (i.e., recA and ATPγS concentration ratios are usually kept constant at a specific concentration of individual polynucleotide strand, depending on whether a single- or double-stranded polynucleotide is used).

RecA protein coating of targeting polynucleotides is normally carried out in a standard 1X RecA coating reaction buffer. 10X RecA reaction buffer (i.e., 10x AC buffer) consists of: 100 mM Tris acetate (pH 7.5 at 37°C), 20 mM magnesium acetate, 500 mM sodium acetate, 10 mM DTT, and 50% glycerol). All of the targeting polynucleotides, whether double-stranded or single-stranded, typically are denatured before use by heating to 95-100°C for five minutes, placed on ice for one minute, and subjected to centrifugation (10,000 rpm) at 0°C for approximately 20 seconds (e.g., in a Tomy centrifuge). Denatured targeting polynucleotides usually are added immediately to room temperature RecA coating reaction buffer mixed with ATPγS and diluted with double-distilled H_2O as necessary.

A reaction mixture typically contains the following components: (i) 0.2-4.8 mM ATPγS; and (ii) between 1-100 ng/μl of targeting polynucleotide. To this mixture is added about 1-20 μl of recA protein per 10-100 μl of reaction mixture, usually at about 2-10 mg/ml (purchased from Pharmacia or purified), and is rapidly added and mixed. The final reaction volume for RecA coating of targeting polynucleotide is usually in the range of about 10-500 μl. RecA coating of targeting polynucleotide is usually initiated by incubating targeting polynucleotide-RecA mixtures at 37°C for about 10-15 min.

RecA protein concentrations in coating reactions varies depending upon targeting polynucleotide size and the amount of added targeting polynucleotide: recA protein concentrations are typically in the range of 5 to 50 μM. When single-stranded targeting polynucleotides are coated with recA, independently of their complementary strands, the concentrations of ATPγS and recA protein may optionally be reduced to about one-half of the concentrations used with double-stranded targeting polynucleotides of the same length: that is, the recA protein and ATPγS concentration ratios are generally kept constant for a given concentration of individual polynucleotide strands.

The coating of targeting polynucleotides with recA protein can be evaluated in a number of ways. First, protein binding to DNA can be examined using band-shift gel assays (McEntee et al., (1981) J. Biol. Chem. 256: 8835). Labeled polynucleotides can be coated with recA protein in the presence of ATPγS and the products of the coating reactions may be separated by agarose gel electrophoresis. Following incubation of recA protein with denatured duplex DNAs the recA protein effectively coats single-stranded targeting polynucleotides derived from denaturing a duplex DNA. As the ratio of recA protein monomers to nucleotides in the targeting polynucleotide increases from 0, 1:27, 1:2.7 to 3.7:1 for 121-mer and 0, 1:22, 1:2.2 to 4.5:1 for 159-mer, targeting polynucleotide's electrophoretic mobility decreases, i.e., is retarded, due to recA-binding to the targeting polynucleotide. Retardation of the coated polynucleotide's mobility reflects the saturation of targeting polynucleotide with recA protein. An excess of recA monomers to DNA nucleotides is required for efficient recA coating of short targeting polynucleotides (Leahy et al., (1986) J. Biol. Chem. 261: 954).

A second method for evaluating protein binding to DNA is in the use of nitrocellulose fiber binding assays (Leahy et al., (1986) J. Biol. Chem. 261:6954; Woodbury, et al., (1983) Biochemistry 22(20):4730-4737. The nitrocellulose filter binding method is particularly useful in determining the dissociation-rates for protein:DNA complexes using labeled DNA. In the filter binding assay, DNA:protein complexes are retained on a filter while free DNA passes through the filter. This assay method is more quantitative for dissociation-rate determinations because the separation of DNA:protein complexes from free targeting polynucleotide is very rapid.

Alternatively, recombinase protein(s) (prokaryotic, eukaryotic or endogeneous to the target cell) may be exogenously induced or administered to a target cell simultaneously or contemporaneously (i.e., within about a few hours) with the targeting polynucleotide(s). Such administration is typically done by micro-injection, although electroporation, lipofection, and other transfection methods known in the art may also be used. Alternatively, recombinase-proteins may be produced in vivo. For example, they may be produced from a homologous or heterologous expression cassette in a transfected cell or transgenic cell, such as a transgenic totipotent cell (e.g. a fertilized zygote) or an embryonal stem cell (e.g., a

murine ES cell such as AB-1) used to generate a transgenic non-human animal line or a somatic cell or a pluripotent hematopoietic stem cell for reconstituting all or part of a particular stem cell population (e.g. hematopoietic) of an individual. Conveniently, a heterologous expression cassette includes a modulatable promoter, such as an ecdysone-inducible promoter-enhancer combination, an estrogen-induced promoter-enhancer combination, a CMV promoter-enhancer, an insulin gene promoter, or other cell-type specific, developmental stage-specific, hormone-inducible, or other modulatable promoter construct so that expression of at least one species of recombinase protein from the cassette can by modulated for transiently producing recombinase(s) in vivo simultaneous or contemporaneous with introduction of a targeting polynucleotide into the cell. When a hormone-inducible promoter-enhancer combination is used, the cell must have the required hormone receptor present, either naturally or as a consequence of expression a co-transfected expression vector encoding such receptor. Alternatively, the recombinase may be endogeneous and produced in high levels. In this embodiment, preferably in eukaryotic target cells such as tumor cells, the target cells produce an elevated level of recombinase. In other embodiments the level of recombinase may be induced by DNA damaging agents, such as mitomycin C, UV or γ-irradiation. Alternatively, recombinase levels may be elevated by transfection of a plasmid encoding the recombinase gene into the cell.

Cell-Uptake Components

A targeting polynucleotide of the invention may optionally be conjugated, typically by covalently or preferably noncovalent binding, to a cell-uptake component. Various methods have been described in the art for targeting DNA to specific cell types. A targeting polynucleotide of the invention can be conjugated to essentially any of several cell-uptake components known in the art. For targeting to hepatocytes, a targeting polynucleotide can be conjugated to an asialoorosomucoid (ASOR)-poly-L-lysine conjugate by methods described in the art and incorporated herein by reference (Wu GY and Wu CH (1987) J. Biol. Chem. 262:4429; Wu GY and Wu CH (1988) Biochemistry 27:887; Wu GY and Wu CH (1988) J. Biol. Chem. 263: 14621; Wu GY and Wu CH (1992) J. Biol. Chem. 267: 12436; Wu et al. (1991) J. Biol. Chem. 266: 14338; and Wilson et al. (1992) J. Biol. Chem. 267: 963, WO92/06180; WO92/05250; and WO91/17761, which are incorporated herein by reference).

Alternatively, a cell-uptake component may be formed by incubating the targeting polynucleotide with at least one lipid species and at least one protein species to form protein-lipid-polynucleotide complexes consisting essentially of the targeting polynucleotide and the lipid-protein cell-uptake component. Lipid vesicles made according to Felgner (W091/17424, incorporated herein by reference) and/or cationic lipidization (WO91/16024, incorporated herein by reference) or other forms for polynucleotide administration (EP 465,529, incorporated herein by reference) may also be employed as cell-uptake components. Nucleases may also be used.

In addition to cell-uptake components, targeting components such as nuclear localization signals may be used, as is known in the art.

Homologous Pairing of Targeting Polynucleotides Having Chemical Substituents

In addition to recombinase and cellular uptake components, the targeting polynucleotides may include chemical substituents. Exogenous targeting polynucleotides that have been modified with appended chemical substituents may be introduced along with recombinase (e.g., recA) into a metabolically active target cell to homologously pair with a predetermined endogenous DNA target sequence in the cell. In a preferred embodiment, the exogenous targeting polynucleotides are derivatized, and additional chemical substituents are attached, either during or after polynucleotide synthesis, respectively, and are thus localized to a specific endogenous target sequence where they produce an alteration or chemical modification to a local DNA sequence. Preferred attached chemical substituents include, but are not limited to: cross-linking agents (see Podyminogin et al., Biochem. 34:13098 (1995) and 35:7267 (1996), both of which are hereby incorporated by reference), nucleic acid cleavage agents, metal chelates (e.g., iron/EDTA chelate for iron catalyzed cleavage), topoisomerases, endonucleases, exonucleases, ligases, phosphodiesterases, photodynamic porphyrins, chemotherapeutic drugs (e.g., adriamycin, doxirubicin), intercalating agents, labels, base-modification agents, agents which normally bind to nucleic acids such as labels, etc. (see for example Afonina et al., PNAS USA 93:3199 (1996), incorporated herein by reference) immunoglobulin chains, and oligonucleotides. Iron/EDTA chelates are particularly preferred chemical substituents where local cleavage of a DNA sequence is

desired (Hertzberg et al. (1982) J. Am. Chem. Soc. 104: 313; Hertzberg and Dervan (1984) Biochemistry 23: 3934; Taylor et al. (1984) Tetrahedron 40: 457; Dervan, PB (1986) Science 232: 464, which are incorporated herein by reference). Further preferred are groups that prevent hybridization of the complementary single stranded nucleic acids to each other but not to unmodified nucleic acids; see for example Kutryavin et al., Biochem. 35:11170 (1996) and Woo et al., Nucleic Acid. Res. 24(13):2470 (1996), both of which are incorporated by reference. 2'-O methyl groups are also preferred; see Cole-Strauss et al., Science 273:1386 (1996); Yoon et al., PNAS 93:2071 (1996)). Additional preferred chemical substitutents include labeling moieties, including fluoroscent labels. Preferred attachment chemistries include: direct linkage, e.g., via an appended reactive amino group (Corey and Schultz (1988) Science 238:1401, which is incorporated herein by reference) and other direct linkage chemistries, although streptavidin/biotin and digoxigenin/antidigoxigenin antibody linkage methods may also be used. Methods for linking chemical substituents are provided in U.S. Patents 5,135,720; 5,093,245, and 5,055,556, which are incorporated herein by reference. Other linkage chemistries may be used at the discretion of the practitioner.

Typically, a targeting polynucleotide of the invention is coated with at least one recombinase and is conjugated to a cell-uptake component, and the resulting cell targeting complex is contacted with a target cell under uptake conditions (e.g., physiological conditions) so that the targeting polynucleotide and the recombinase(s) are internalized in the target cell. A targeting polynucleotide may be contacted simultaneously or sequentially with a cell-uptake component and also with a recombinase; preferably the targeting polynucleotide is contacted first with a recombinase, or with a mixture comprising both a cell-uptake component and a recombinase under conditions whereby, on average, at least about one molecule of recombinase is noncovalently attached per targeting polynucleotide molecule and at least about one cell-uptake component also is noncovalently attached. Most preferably, coating of both recombinase and cell-uptake component saturates essentially all of the available binding sites on the targeting polynucleotide. A targeting polynucleotide may be preferentially coated with a cell-uptake component so that the resultant targeting complex comprises, on a molar basis, more cell-uptake component than recombinase(s). Alternatively, a targeting polynucleotide may be preferentially coated with recombinase(s) so that the resultant

targeting complex comprises, on a molar basis, more recombinase(s) than cell-uptake component.

Cell-uptake components are included with recombinase-coated targeting polynucleotides of the invention to enhance the uptake of the recombinase-coated targeting polynucleotide(s) into cells, particularly for in vivo gene targeting applications, such as gene therapy to treat genetic diseases, including neoplasia, and targeted homologous recombination to treat viral infections wherein a viral sequence (e.g., an integrated hepatitis B virus (HBV) genome or genome fragment) may be targeted by homologous sequence targeting and inactivated. Alternatively, a targeting polynucleotide may be coated with the cell-uptake component and targeted to cells with a contemporaneous or simultaneous administration of a recombinase (e.g., liposomes or immunoliposomes containing a recombinase, a viral-based vector encoding and expressing a recombinase).

Once the recombinase-targeting polynucleotide compositions are formulated, they are introduced or administered into target cells. The administration is typically done as is known for the administration of nucleic acids into cells, and, as those skilled in the art will appreciate, the methods may depend on the choice of the target cell. Suitable methods include, but are not limited to, microinjection, electroporation, lipofection, etc. By "target cells" herein is meant prokaryotic or eukaryotic cells. Suitable prokaryotic cells include, but are not limited to, bacteria such as *E. coli*, *Bacillus* species, and the extremophile bacteria such as thermophiles, etc. Preferably, the procaryotic target cells are recombination competent. Suitable eukaryotic cells include, but are not limited to, fungi such as yeast and filamentous fungi, including species of *Aspergillus*, *Trichoderma*, and *Neurospora*; plant cells including those of corn, sorghum, tobacco, canola, soybean, cotton, tomato, potato, alfalfa, sunflower, etc.; and animal cells, including fish, birds and mammals. Suitable fish cells include, but are not limited to, those from species of salmon, trout, tulapia, tuna, carp, flounder, halobut, swordfish, cod and zebrafish. Suitable bird cells include, but are not limited to, those of chickens, ducks, quail, pheasants and turkeys, and other jungle foul or game birds. Suitable mammalian cells include, but are not limited to, cells from horses, cows, buffalo, deer, sheep, rabbits, rodents such as mice, rats, hamsters and guinea pigs,

goats, pigs, primates, marine mammals including dolphins and whales, as well as cell lines, such as human cell lines of any tissue or stem cell type, and stem cells, including pluripotent and non-pluripotent, and non-human zygotes.

In a preferred embodiment, procaryotic cells are used. In this embodiment, a pre-selected target DNA sequence is chosen for alteration. Preferably, the pre-selected target DNA sequence is contained within an extrachromosomal sequence. By "extrachromosomal sequence" herein is meant a sequence separate from the chromosomal or genomic sequences. Preferred extrachromosomal sequences include plasmids (particularly procaryotic plasmids such as bacterial plasmids), p1 vectors, viral genomes, yeast, bacterial and mammalian artificial chromosomes (YAC, BAC and MAC, respectively), and other autonomously self-replicating sequences, although this is not required. As described herein, a recombinase and at least two single stranded targeting polynucleotides which are substantially complementary to each other, each of which contain a homology clamp to the target sequence contained on the extrachromosomal sequence, are added to the extrachromosomal sequence, preferably *in vitro*. The two single stranded targeting polynucleotides are preferably coated with recombinase, and at least one of the targeting polynucleotides contain at least one nucleotide substitution, insertion or deletion. The targeting polynucleotides then bind to the target sequence in the extrachromosomal sequence to effect homologous recombination and form an altered extrachromosomal sequence which contains the substitution, insertion or deletion. The altered extrachromosomal sequence is then introduced into the procaryotic cell using techniques known in the art. Preferably, the recombinase is removed prior to introduction into the target cell, using techniques known in the art. For example, the reaction may be treated with proteases such as proteinase K, detergents such as SDS, and phenol extraction (including phenol:chloroform:isoamyl alcohol extraction). These methods may also be used for eukaryotic cells.

Alternatively, the pre-selected target DNA sequence is a chromosomal sequence. In this embodiment, the recombinase with the targeting polynucleotides are introduced into the target cell, preferably eukaryotic target cells. In this embodiment, it may be desirable to bind (generally non-covalently) a nuclear localization signal to the targeting polynucleotides to

facilitate localization of the complexes in the nucleus. See for example Kido et al., Exper. Cell Res. 198:107-114 (1992), hereby expressly incorporated by reference. The targeting polynucleotides and the recombinase function to effect homologous recombination, resulting in altered chromosomal or genomic sequences.

In a preferred embodiment, eukaryotic cells are used. For making transgenic non-human animals (which include homologously targeted non-human animals) embryonal stem cells (ES cells) and fertilized zygotes are preferred. In a preferred embodiment, embryonal stem cells are used. Murine ES cells, such as AB-1 line grown on mitotically inactive SNL76/7 cell feeder layers (McMahon and Bradley, Cell 62: 1073-1085 (1990)) essentially as described (Robertson, E.J. (1987) in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach. E.J. Robertson, ed. (oxford: IRL Press), p. 71-112) may be used for homologous gene targeting. Other suitable ES lines include, but are not limited to, the E14 line (Hooper et al. (1987) Nature 326: 292-295), the D3 line (Doetschman et al. (1985) J. Embryol. Exp. Morph. 87: 21-45), and the CCE line (Robertson et al. (1986) Nature 323: 445-448). The success of generating a mouse line from ES cells bearing a specific targeted mutation depends on the pluripotence of the ES cells (i.e., their ability, once injected into a host blastocyst, to participate in embryogenesis and contribute to the germ cells of the resulting animal).

The pluripotence of any given ES cell line can vary with time in culture and the care with which it has been handled. The only definitive assay for pluripotence is to determine whether the specific population of ES cells to be used for targeting can give rise to chimeras capable of germline transmission of the ES genome. For this reason, prior to gene targeting, a portion of the parental population of AB-1 cells is injected into C57B1/6J blastocysts to ascertain whether the cells are capable of generating chimeric mice with extensive ES cell contribution and whether the majority of these chimeras can transmit the ES genome to progeny.

In a preferred embodiment, non-human zygotes are used, for example to make transgenic animals, using techniques known in the art (see U.S. Patent No. 4,873,191). Preferred zygotes include, but are not limited to, animal zygotes, including fish, avian and mammalian

zygotes. Suitable fish zygotes include, but are not limited to, those from species of salmon, trout, tuna, carp, flounder, halibut, swordfish, cod, tulapia and zebrafish. Suitable bird zygotes include, but are not limited to, those of chickens, ducks, quail, pheasant, turkeys, and other jungle fowl and game birds. Suitable mammalian zygotes include, but are not limited to, cells from horses, cows, buffalo, deer, sheep, rabbits, rodents such as mice, rats, hamsters and guinea pigs, goats, pigs, primates, and marine mammals including dolphins and whales. See Hogan et al., Manipulating the Mouse Embryo (A Laboratory Manual), 2nd Ed. Cold Spring Harbor Press, 1994, incorporated by reference.

The vectors containing the DNA segments of interest can be transferred into the host cell by well-known methods, depending on the type of cellular host. For example, micro-injection is commonly utilized for target cells, although calcium phosphate treatment, electroporation, lipofection, biolistics or viral-based transfection also may be used. Other methods used to transform mammalian cells include the use of Polybrene, protoplast fusion, and others (see, generally, Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed., 1989, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference). Direct injection of DNA and/or recombinase-coated targeting polynucleotides into target cells, such as skeletal or muscle cells also may be used (Wolff et al. (1990) Science 247: 1465, which is incorporated herein by reference).

Targeting of Endogenous DNA Sequences

Once made and administered to a target host cell, the compositions of the invention find use in a number of applications, including the site directed modification of endogeneous sequences within any target cell, the creation of transgenic plants and animals, and the use of the compositions to do site-directed mutagenesis or modifications of target sequences.

Generally, any predetermined endogenous DNA sequence, such as a gene sequence, can be altered by homologous recombination (which includes gene conversion) with an exogenous targeting polynucleotides (such as a complementary pair of single-stranded targeting polynucleotides). The target polynucleotides have at least one homology clamp which substantially corresponds to or is substantially complementary to a predetermined

endogenous DNA target sequence and are introduced with a recombinase (e.g., recA) into a target cell having the predetermined endogenous DNA sequence. Typically, a targeting polynucleotide (or complementary polynucleotide pair) has a portion or region having a sequence that is not present in the preselected endogenous targeted sequence(s) (i.e., a nonhomologous portion or mismatch) which may be as small as a single mismatched nucleotide, several mismatches, or may span up to about several kilobases or more of nonhomologous sequence. Generally, such nonhomologous portions are flanked on each side by homology clamps, although a single flanking homology clamp may be used. Nonhomologous portions are used to make insertions, deletions, and/or replacements in a predetermined endogenous targeted DNA sequence, and/or to make single or multiple nucleotide substitutions in a predetermined endogenous target DNA sequence so that the resultant recombined sequence (i.e., a targeted recombinant endogenous sequence) incorporates some or all of the sequence information of the nonhomologous portion of the targeting polynucleotide(s). Thus, the nonhomologous regions are used to make variant sequences, i.e. targeted sequence modifications. Additions and deletions may be as small as 1 nucleotide or may range up to about 2 to 4 kilobases or more. In this way, site directed directed modifications may be done in a variety of systems for a variety of purposes.

In a preferred application, a targeting polynucleotide is used to repair a mutated sequence of a structural gene by replacing it or converting it to a wild-type sequence (e.g., a sequence encoding a protein with a wild-type biological activity). For example, such applications could be used to convert a sickle cell trait allele of a hemoglobin gene to an allele which encodes a hemoglobin molecule that is not susceptible to sickling, by altering the nucleotide sequence encoding the β-subunit of hemoglobin so that the codon at position 6 of the β-subunit is converted Valβ6-->Gluβ6 (Shesely et al. (1991) op.cit.). Other genetic diseases can be corrected, either partially or totally, by replacing, inserting, and/or deleting sequence information in a disease allele using appropriately selected exogenous targeting polynucleotides. For example but not for limitation, the ΔF508 deletion in the human CFTR gene can be corrected by targeted homologous recombination employing a recA-coated targeting polynucleotide of the invention.

For many types of in vivo gene therapy to be effective, a significant number of cells must be correctly targeted, with a minimum number of cells having an incorrectly targeted recombination event. To accomplish this objective, the combination of: (1) a targeting polynucleotide(s), (2) a recombinase (to provide enhanced efficiency and specificity of correct homologous sequence targeting), and (3) a cell-uptake component (to provide enhanced cellular uptake of the targeting polynucleotide), provides a means for the efficient and specific targeting of cells in vivo, making in vivo homologous sequence targeting, and gene therapy, practicable.

Several disease states may be amenable to treatment or prophylaxis by targeted alteration of heptocytes in vivo by homologous gene targeting. For example and not for limitation, the following diseases, among others not listed, are expected to be amenable to targeted gene therapy: hepatocellular carcinoma, HBV infection, familial hypercholesterolemia (LDL receptor defect), alcohol sensitivity (alcohol dehydrogenase and/or aldehyde dehydrogenase insufficiency), hepatoblastoma, Wilson's disease, congenital hepatic porphyrias, inherited disorders of hepatic metabolism, ornithine transcarbamylase (OTC) alleles, HPRT alleles associated with Lesch Nyhan syndrome, etc. Where targeting of hepatic cells in vivo is desired, a cell-uptake component consisting essentially of an asialoglycoprotein-poly-L-lysine conjugate is preferred. The targeting complexes of the invention which may be used to target hepatocytes in vivo take advantage of the significantly increased targeting efficiency produced by association of a targeting polynucleotide with a recombinase which, when combined with a cell-targeting method such as that of WO92/05250 and/or Wilson et al. (1992) J. Biol. Chem. 267:963, provide a highly efficient method for performing in vivo homologous sequence targeting in cells, such as hepatocytes.

In a preferred embodiment, the methods and compositions of the invention are used for gene inactivation. That is, in addition to correcting disease alleles, exogenous targeting polynucleotides can be used to inactivate, decrease or alter the biological activity of one or more genes in a cell (or transgenic nonhuman animal). This finds particular use in the generation of animal models of disease states, or in the elucidation of gene function and activity, similar to "knock out" experiments. These techniques may be used to eliminate a

biological function; for example, a galT gene (alpha galactosyl transferase genes) associated with the xenoreactivity of animal tissues in humans may be disrupted to form transgenic animals (e.g. pigs) to serve as organ transplantation sources without associated hyperacute rejection responses. Alternatively, the biological activity of the wild-type gene may be either decreased, or the wild-type activity altered to mimic disease states. This includes genetic manipulation of non-coding gene sequences that affect the transcription of genes, including, promoters, repressors, enhancers and transcriptional activating sequences.

Once the specific target genes to be modified are selected, their sequences may be scanned for possible disruption sites (convenient restriction sites, for example). Plasmids are engineered to contain an appropriately sized gene sequence with a deletion or insertion in the gene of interest and at least one flanking homology clamp which substantially corresponds or is substantially complementary to an endogenous target DNA sequence. Vectors containing a targeting polynucleotide sequence are typically grown in *E. coli* and then isolated using standard molecular biology methods, or may be synthesized as oligonucleotides. Direct targeted inactivation which does not require vectors may also be done. When using microinjection procedures it may be preferable to use a transfection technique with linearized sequences containing only modified target gene sequence and without vector or selectable sequences. The modified gene site is such that a homologous recombinant between the exogenous targeting polynucleotide and the endogenous DNA target sequence can be identified by using carefully chosen primers and PCR, followed by analysis to detect if PCR products specific to the desired targeted event are present (Erlich et al., (1991) Science 252: 1643, which is incorporated herein by reference). Several studies have already used PCR to successfully identify and then clone the desired transfected cell lines (Zimmer and Gruss, (1989) Nature 338: 150; Mouellic et al., (1990) Proc. Natl. Acad. Sci. USA 87: 4712; Shesely et al., (1991) Proc. Natl. Acad. Sci. USA 88: 4294, which are incorporated herein by reference). This approach is very effective when the number of cells receiving exogenous targeting polynucleotide(s) is high (i.e., with microinjection, or with liposomes) and the treated cell populations are allowed to expand to cell groups of approximately 1 x 10^4 cells (Capecchi, (1989) Science 244: 1288). When the target gene is not on a sex chromosome, or

labelled slower-migrating species provides further evidence for the existence of the multi-stranded DNA hybrids.

EcoR1 Restriction Endonucleases cut duplex DNA in either homologous or heterologous cssDNA probe:target hybrids. To further characterize cssDNA probe:target hybrids formed with heterologous DNA targets, circular plasmids pRD.0 and pRD.59 were hybridized with biotin-labelled probe IP290 and then deproteinized and digested with EcoRI. While plasmid pRD.0 contains a unique EcoR1 site in the region of homology between IP290 and pRD.0, the EcoR1 site is deleted in pRD.59 (Figure 14A). Digestion of cssDNA probe:target hybrids with EcoR1 indicates the restoration of Watson-Crick pairing to form a fully duplex EcoR1 recognition site. Figure 18 shows both the ethidium bromide stained gel of the hybrid product of the targeting reaction (Figure 18A and 18B) and the corresponding autoradiograph that shows the electrophoretic migration of the biotin-labelled probes (Figure 18C and 18D). These data show that when RecA-coated IP290 is hybridized to the fully homologous pRD.0 plasmid all of the probe:target hybrids migrate to position of fully relaxed DNA (Figure 18 A and C, Lane 1). Furthermore, upon digestion with EcoR1 cssDNA:probe target hybrids can be completely cut as shown in Figure 18 A and C, Lane 2. When similar reactions are performed with uncut pRD.59 targets, we found that not all of the probe:target hybrids are relaxed as with pRD.0 targets, as judged by the appearance of two bands corresponding to a pRD59 I* hybrid, where the hybrids co-migrate with FormI supercoiled DNA and a pRD59 rI* hybrid that migrates with relaxed targets (Figure 18B and D, Lane 3). When these hybrids are digested with EcoRI we find that the pRD59 rI* hybrid is more susceptible to EcoRI cleavage than the pRD59 rI* hybrid (Figure 18B and D, Lane 4). This shows that there is a restoration of the EcoRI site in relaxed targets, but not in the non-relaxed I* hybrid. Since pRD59 targets do not contain an EcoRI site, cleavage by EcoRI can only be explained by re-annealing of cssDNA probe IP290 within the IP290 probe:target pRD59 hybrid.

To further characterize the structural differences between pRD59 rI* hybrids and pRD59 I* hybrids, cssDNA probe:target hybrids were formed between IP290 and pRD59, deproteinized and thermally melted for 5 mins at 37°C, 45°C, 55°C, and 65°C, respectively. Figure 19 shows that pRD59 rI* hybrids are more thermostable than pRD59 I* hybrids. For both types

of hybrids probe:target hybrids are completely dissociated after heating to 95°C (data not shown). Taken together these data support the structures of our models for hybrids (Figure 13).

EXAMPLE 6

Homologous recombination targeting in fertilized mouse zygotes

Ornithine transcarbamylase (OTC) is a mitochondrial matrix enzyme that catalyzes the synthesis of citrulline from ornithine and carbamylphosphate in the second step of the mammalian urea cycle. OTC deficiency in humans is the most common and severe defect of the urea cycle disorders. OTC is an X-linked gene that is primarily expressed in the liver and to a lesser extent in the small intestine. Affected males develop hyperammonemia, acidosis, orotic aciduria, coma and death occurs in up to 75% of affected males, regardless of intervention. Two allelic mutations at the OTC locus are known in mice: spf and spf-ash, (sparse fur--abnormal skin and hair). In addition to hyperammonemia and orotic aciduria spf-ash mice can be readily identified by the abnormal skin and hair phenotype. The spf-ash mutation is a single-base substitution at the end of exon 4 that results in alternative intron-exon splicing to produce of an aberrant non-functional elongated pre-mRNA. Because of the clinical importance of OTC defects in humans, there is an intensive effort to develop in vivo methods to correct the enzymatic defect in the spf ash mouse model.

We used the spf-ash murine model of OTC deficiency to test the ability of RecA-coated complementary single-stranded DNA (css) OTC probes to target and correct a single-base substitution mutation in fertilized mouse zygotes. A 230 bp RecA-coated cssDNA probe amplified from the normal mouse OTC gene was microinjected into embryos made from the cross of B6C3H homozygous female spf-ash and normal B6D2F1J males. After re-implantation of 75 embryos that were microinjected with RecA-coated cssDNA into CD1 foster mothers, 25 developmentally normal pups (17 female and 8 male) were produced. Sequence analysis of the genomic DNA isolated from tails of the male pups show that in 3/8 males a homologous recombination event occured that produced mosaic animals at the spf-ash site in exon4 of the mouse OTC gene. Subsequent breeding of the three the mosaic

male founder mice with normal females demonstrated the gene corrected OTC allele was transmitted to the sperm germline from one of these three mosaic homologous recombinant mice, as determined by sequence analysis of the genomic DNA and transmission of phenotypic correction to F1 mice. These studies illustrate the utility of cssDNA probes to mediate high frequency homologous recombination in fertilized mouse zygotes to create subtle genetic modifications at a desired target site in the chromosome.

Preparation of RecA-coated probe: A 230 bp fragment from the normal mouse OTC gene was amplified by PCR with primers M9 and M8 from pTAOTC (Figure 20). The PCR fragment was purified on Microcon-100 columns (Amicon) and then extensively dialyzed in ddH_2O. The M9-M8 amplicon was denatured by heating the fragments to 98°C and then coated with RecA protein (Boehringer-Mannheim) at a ratio 3 nucleotides/ protein monomer. The final concentration of RecA-coated DNA in coating buffer (5 mM TrisOAc, pH 7.5, 0.5 mM DTT, 10 mM MgOAc, 1.22 mM ATPγS, 5.5 μM RecA) was 5 ng/ μL. RecA-coated filaments were made on the day of microinjection and then stored on ice until use.

Transgenic Mice: Five superovulated B6C3H (spf-ash/spf-ash) 5-7 week old females (Jackson Labs) were mated with five B6D2F1 males (Jackson Labs). Approximately 80-100 embryos were isolated from oviducts as described in Hogan et al. (1988). The female pronucleus of fertilized embryos were microinjected with 2 pl of RecA-coated M9-M8 cssDNA probe (5 ng/μL). Approximately 75 embryos survived the microinjection procedure and were then re-implanted into a total of three CD1 pseudopregnant foster mothers (Charles River). Pseudopregnant females were produced by mating foster mothers with vasectomized CD1 males (Charles River).

DNA Analysis: Tail biopsies were taken from all founder mice after weaning at and ear-tagging at three weeks of age. Genomic DNA was isolated from tail biopsies using standard procedures. To obtain the sequence of the DNA at the OTC locus, genomic DNA was amplified with PCR using primers M10-M11 or M54-M11 that flank the cssDNA probe sequence to generate a 250 bp or 314 bp amplicon (Figure 20). PCR fragments were sequenced manually using the Cyclist Exo Kit (Stratagene), automatically on Applied

Biosystems Model 373A sequencer, or by a MALDI-TOF mass spectrometry system (GeneTrace Systems, Menlo Park, CA)

Fertilized zygotes microinjected with RecA-coated DNA are viable. Plasmid pTAOTC1 carries a 250 bp segment of exon4 and surrounding intron sequences from the normal mouse OTC gene. A 230 bp cssDNA probe OTC1 was prepared by PCR amplification of pTAOTC1 with primers M9 and M8. cssDNA probe OTC1 was denatured and coated with RecA protein as described herein.

Homozygous spf-ash/spf-ash female and hemizygous (spf-ash/y) males can be phenotypically identified by the appearance of sparse fur and wrinkled skin early in development. A cross between homozygous spf-ash/spf-ash B6C3H females and normal B6D2F1 males yields heterozygous phenotypically normal females and hemizygous males with sparse fur and wrinkled skin. The RecA-coated cssDNA OTC probe was microinjected into embryos made from the cross of B6C3H homozygous female spf-ash and normal males. The female pronucleus of approximately 80-90 fertilized zygotes was microinjected with 2 pl of a 5ng/μL solution of RecA-coated cssDNA probe OTC1. Of these 75 embryos survived the microinjection procedure. To demonstrate that embryos that have been microinjected with RecA-coated cssDNA are viable, the embryos were re-implanted into three pseudopregnant CD1 foster mothers. From this, 25 developmentally normal pups (17 female and 8 male) were produced. All of the female mice were phenotypically normal. The eight male mice (mouse # 7, 14,16,17,22,23,24, and 25) were all affected with sparse-fur and wrinkled skin to various degrees.

RecA-coated cssDNA probe OTC1 recombines with the homologous chromosomal copy of the OTC gene in fertilized mouse zygotes. To determine the genotypes of the 25 founder mice produced from microinjected embryos, genomic DNA was isolated from tail biopsies containing skin, blood and bone cells. Genomic DNA was amplified with either the primer set M10-M11 or M54-M11 to produce either a 250 bp or 314 bp amplicon. By using these primer sets that flank the OTC1 probe, the DNA amplicon represents DNA from the endogenous OTC gene. PCR fragments from all of the eight mice and several female mice

were sequenced to determine the base sequence at the spf-ash locus to determine if a normal allele (G) or a mutant allele (A) was present in the genomic DNA. Figure 21 shows sequencing gels of representative reactions. The leftmost panel shows the sequence of the homozygous spf-ash females that donated the eggs to produce the fertilized zygotes where only the mutant base A is present at the spf-ash locus, as expected. The sequence of female mouse #8 that should be heterozygous shows the presence of equal amounts of the bases G and A as expected. Male mice 7, 14 (shown), 23, 24,and 25 all showed only the mutant base A at the spf-ash locus, however male mice 16, 17, and 22 (shown) displayed both G (normal) and A (mutant) at the spf-ash locus.

To eliminate the possibility of PCR artifacts during PCR cycle sequencing the base compositions of the samples was independently confirmed by mass spectrometry sequencing (GeneTrace, Menlo Park). The relative amounts of the A:G base composition at the spf-ash locus was also quantified and determined to be 70:30 for samples from mouse #16 and #17 and 10:90 for mouse#22. Since OTC is an X-linked gene the presence of mixed bases in male mice is likely the result of the mosaic animals produced of a mixture of mutant and gene corrected embryonic cells.

Germline transmission of the gene corrected OTC allele. To determine if the gene corrected allele in the mosaic male founder mice 16, 17, and 22 could be passed through to the germline, these mice and a control hemizygous mutant male #7 were bred with normal B6D2F1 females. In this cross if the male donates a mutant spf-ash X chromosome the resulting female progeny will be heterozygous spf-ash mutants. However if the male donates a normal (gene corrected) X chromosome the female progeny will be homozygous normal. In both cases the resulting F1 females will be phenotypically normal. The results of these crosses are summarized in Figure 22. In the control cross of hemizygous mutant male#7 with B6D2F1 females, all 14 female progeny were heterozygous, as expected. In test crosses of mosaic male mouse #17 and #22 with normal females all resulting female progeny (5 and 9, respectively) were heterozygous. However in the cross with mosaic male mouse #16, one out nine total female progeny was a homozygous normal female (mouse # 213) as determined

mass spectrometry sequencing (GeneTrace, Menlo Park), demonstrating the gene corrected allele in founder mouse #16 was transmitted through the germline.

To further verify that F1 mouse #213 was in fact a germline-transmitted gene corrected homozygous normal female, this and a control heterozygous spf-ash/X mouse were bred with normal males. In the control cross B with the heterozygous female, 50% of the resulting male F2 progeny should be mutant spf-ash/y hemizygotes that can be easily determined by the visualization of sparse-fur and wrinkled skin. Of the 38 progeny produced in this control cross B, 14 were male, and of these, 8 were phenotypically normal and 6 were mutant as determined by the presence of wrinkled skin and abnormal fur. In the test cross with F1 mouse #213, of the 35 progeny produced in this cross, all eleven of the male progeny were phenotypically normal, clearly showing the genotyping of F1 mouse #213 as a germline transmitted gene corrected homozygous normal female.

As another test to determine if the normal gene corrected allele in mouse #16 could be transmitted through the germline, mouse #16 was mated with homozygous (spf-ash/spf-ash) mutant females. In this cross if mouse #16 does not transmit a normal allele, the resultant progeny will either be hemizygous (spf-ash/Y) mutant males or homozygous (spf-ash/spf-ash) mutant females, both of which are phenotypically mutant. However if the mouse allele is transmitted through the germline, heterozygous (spf-ash/+) females that are phenotypically normal will be produced. When mouse #16 was bred with homozygous (spf-ash/spf-ash) mutant females, two litters were produced that consisted of a total 5 hemizygous (spf-ash/Y) mutant males, 7 homozygous (spf-ash/spf-ash) mutant females, and 1 phenotypically normal female (mouse #1014). Pictures of representative mice from these crosses are shown in Figure 23. The production of the phenotypically normal female mouse provides compelling genetic evidence that mouse#16 contains a normal gene corrected OTC allele that is germline transmissable.

Although the present invention has been described in some detail by way of illustration for purposes of clarity of understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the claims.

CLAIMS

We claim:

1. A method for making a targeted sequence modification in a preselected target DNA sequence in a eukaryotic zygote by homologous recombination, said method comprising introducing into at least one eukaryotic zygote at least one recombinase and at least two single-stranded targeting polynucleotides that are substantially complementary to each other and each having a homology clamp that substantially corresponds to or is substantially complementary to a preselected target DNA sequence.

2. A method for making a targeted sequence modification in a preselected target DNA sequence in a cell by homologous recombination, said modification comprising an insertion, said method comprising introducing into at least one cell at least one recombinase and at least two single-stranded targeting polynucleotides which are substantially complementary to each other, each having a homology clamp that substantially corresponds to or is substantially complementary to a preselected target DNA sequence, and each having an internal homology clamp.

3. A method for targeting and altering, by homologous recombination, a pre-selected target nucleic acid sequence in a procaryotic cell to make a targeted sequence modification, said method comprising introducing into at least one procaryotic cell at least one recombinase and at least two single-stranded targeting polynucleotides each of which are substantially complementary to each other and comprise a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence.

4. A method for targeting and altering, by homologous recombination, a pre-selected target nucleic acid sequence in an extrachromosomal sequence of a procaryotic cell, said method comprising:

a) adding to said extrachromosomal sequence at least one recombinase and at least two single-stranded targeting polynucleotides each of which are substantially complementary to each other and comprise a homology clamp

that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence to form an altered extrachromosomal sequence;

b) removing said recombinase; and

c) introducing said altered element into a procaryotic cell.

5. A method of generating a pool of variant nucleic acid sequences of a pre-selected target nucleic acid sequence in an extrachromosomal sequence, said method comprising adding to said extrachromosomal sequence at least one recombinase and a plurality of pairs of single-stranded targeting polynucleotides which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence, said plurality of pairs comprising a library of mismatches between said targeting polynucleotide and said target nucleic acid sequence, to form a library of altered extrachromosomal sequences.

6. A method of generating a cellular library comprising variant nucleic acid sequences of a pre-selected target nucleic acid sequence, said method comprising introducing into a population of target cells at least one recombinase and a plurality of pairs of single-stranded targeting polynucleotides which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence, said plurality of pairs comprising a library of mismatches between said targeting polynucleotide and said target nucleic acid sequence, to form said cellular library comprising variant nucleic acid sequences.

7. A method of generating a cellular library comprising variant nucleic acid sequences of a pre-selected target nucleic acid sequence in an extrachromosomal sequence of a target cell, said method comprising:

a) adding to said extrachromosomal sequence at least one recombinase and a plurality of pairs of single-stranded targeting polynucleotides which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a

preselected target nucleic acid sequence, said plurality of pairs comprising a library of mismatches between said targeting polynucleotide and said target nucleic acid sequence, to form a plurality of altered extrachromosomal sequences;

b) removing said recombinase; and

c) introducing said altered sequences into a population of target cells to form said library of variant nucleic acid sequences.

8. A method according to claim 5 further comprising transforming said pool of altered sequences into a population of cells.

9. A method according to claim 8 wherein said cells are procaryotic.

10. A method according to claim 6 or 7 wherein said cells are eukaryotic.

11. A method according to any of claims 1 to10 further comprising identifying a target cell having a targeted DNA sequence modification at a preselected target DNA sequence.

12. A method according to any of claims 1 to 11, wherein said targeting polynucleotides are coated with said recombinase.

13. A method according to any of claims 1 to 12 wherein said recombinase is a species of prokaryotic recombinase.

14. A method according to claim 13 wherein said prokaryotic recombinase is a species of prokaryotic recA protein.

15. A method according to claim 14 wherein said recA protein species is *E. coli* recA.

16. A method according to any of claims 1 to 12 wherein said recombinase is a species of eukaryotic recombinase.

17. A method according to claim 16, wherein said recombinase is a recombinase from the Rad52 epistasis group.

18. A method according to claim 17, wherein said eukaryotic recombinase is a complex of recombinase proteins.

19. A method according to any of claims 1 to 18 wherein said targeting polynucleotide is conjugated to a cell-uptake component.

20. A method according to any of claims 1 and 3 to 19 wherein the targeted sequence modification comprises the substitution of at least one nucleotide.

21. A method according to claim 20, wherein the targeted sequence modification comprises a plurality of substitutions.

22. A composition comprising at least one recombinase and a variant library comprising a plurality of pairs of single stranded targeting polynucleotides which are substantially complementary to each other and each comprising a homology clamp that substantially corresponds to or is substantially complementary to a preselected target nucleic acid sequence, said plurality of pairs comprising a library of mismatches between said targeting polynucleotide and said target nucleic acid sequence.

23. A composition according to claim 22 wherein said targeting polynucleotides are coated with recombinase.

24. A composition according to claim 22 wherein said recombinase is a species of prokaryotic recombinase.

25. A kit comprising the composition of claim 22 and at least one reagent.

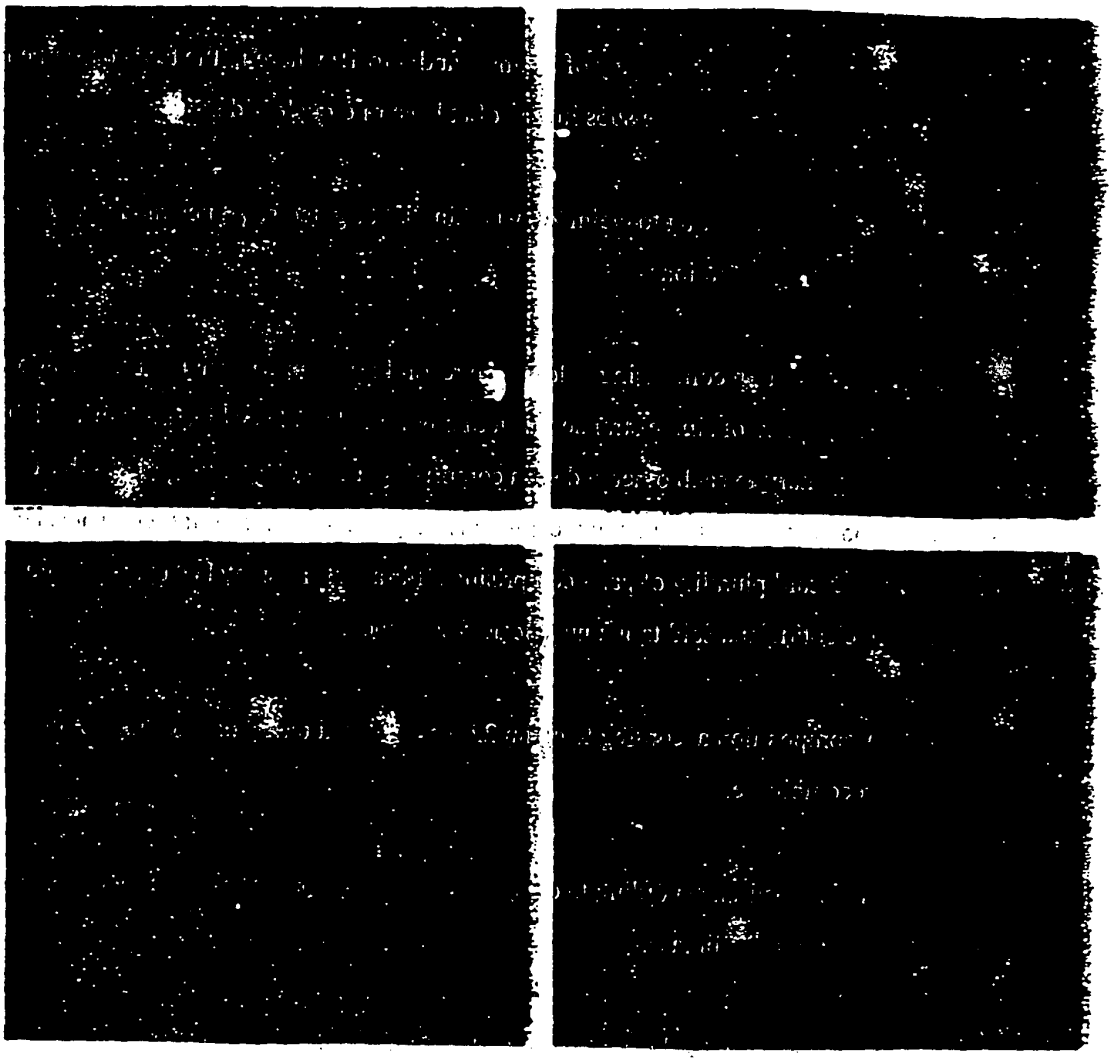


Figure 1

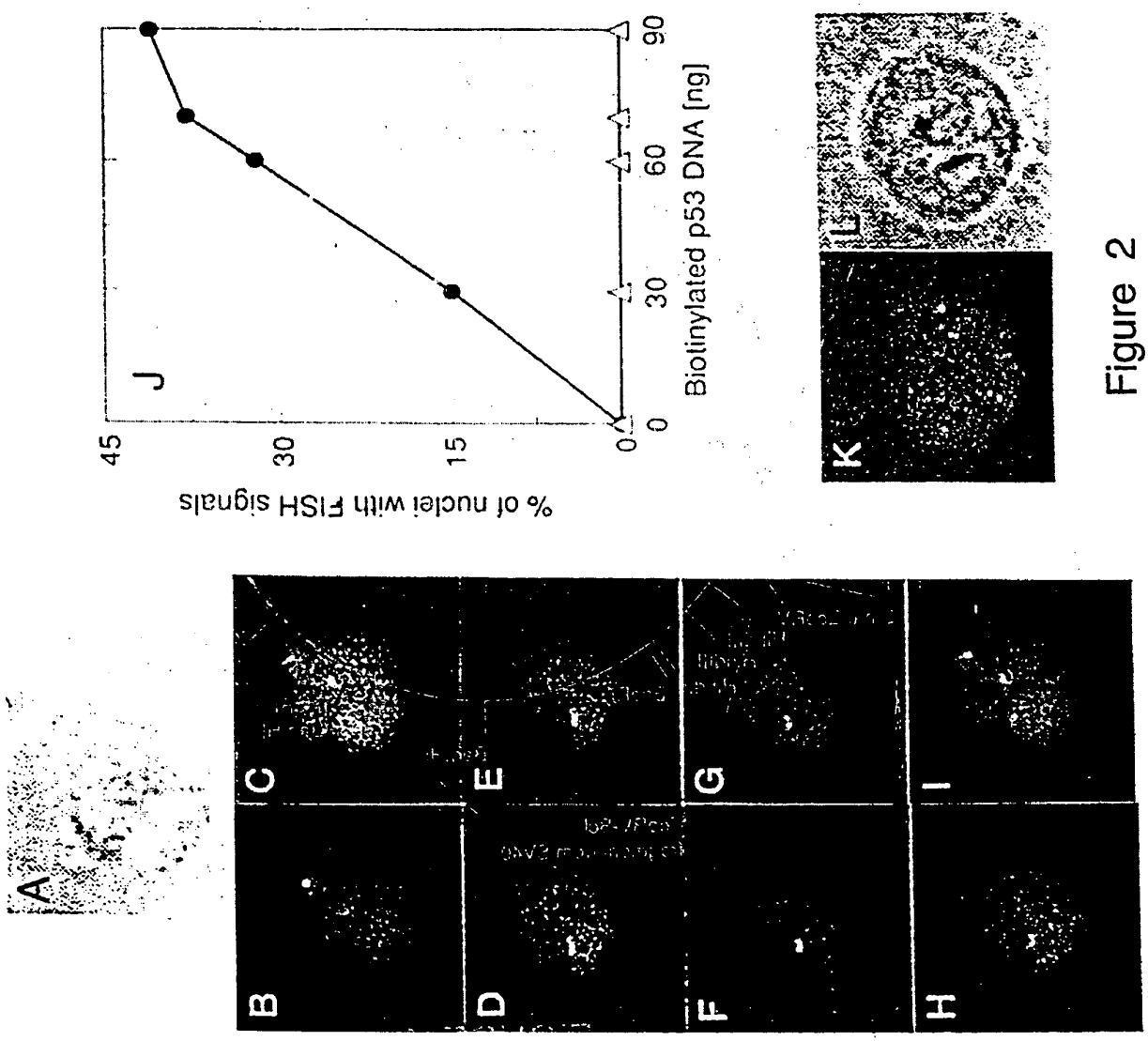

A
B
C
D
E
F
G
H
I
J
K
L
% of nuclei with FISH signals
45
30
15
0
0
30
60
90
Biotinylated p53 DNA [ng]

Figure 2

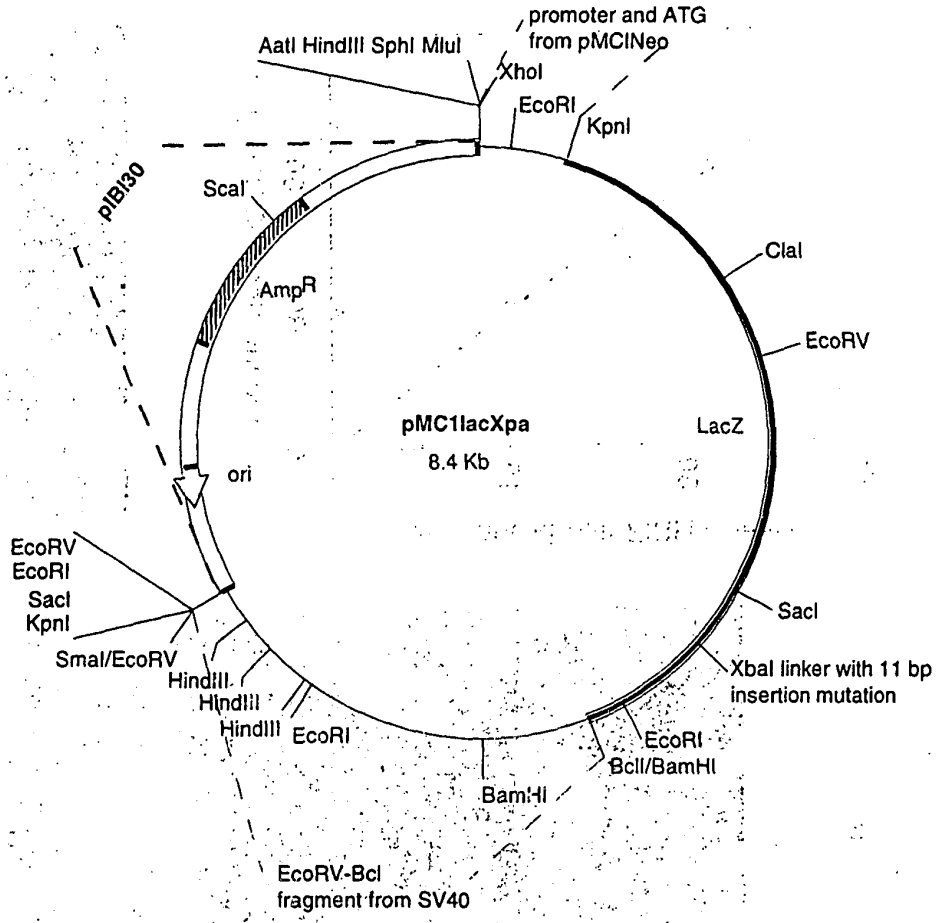

promoter and ATG
from pMClNeo
AatI HindIII SphI MluI
XhoI
EcoRI
KpnI
pIBI30
ScaI
ClaI
AmpR
EcoRV
pMC1lacXpa
8.4 Kb
LacZ
ori
EcoRV
EcoRI
SacI
KpnI
SmaI/EcoRV
HindIII
HindIII
HindIII
EcoRI
SacI
XbaI linker with 11 bp
insertion mutation
EcoRI
BclI/BamHI
BamHI
EcoRV-BclI
fragment from SV40

FIGURE 3

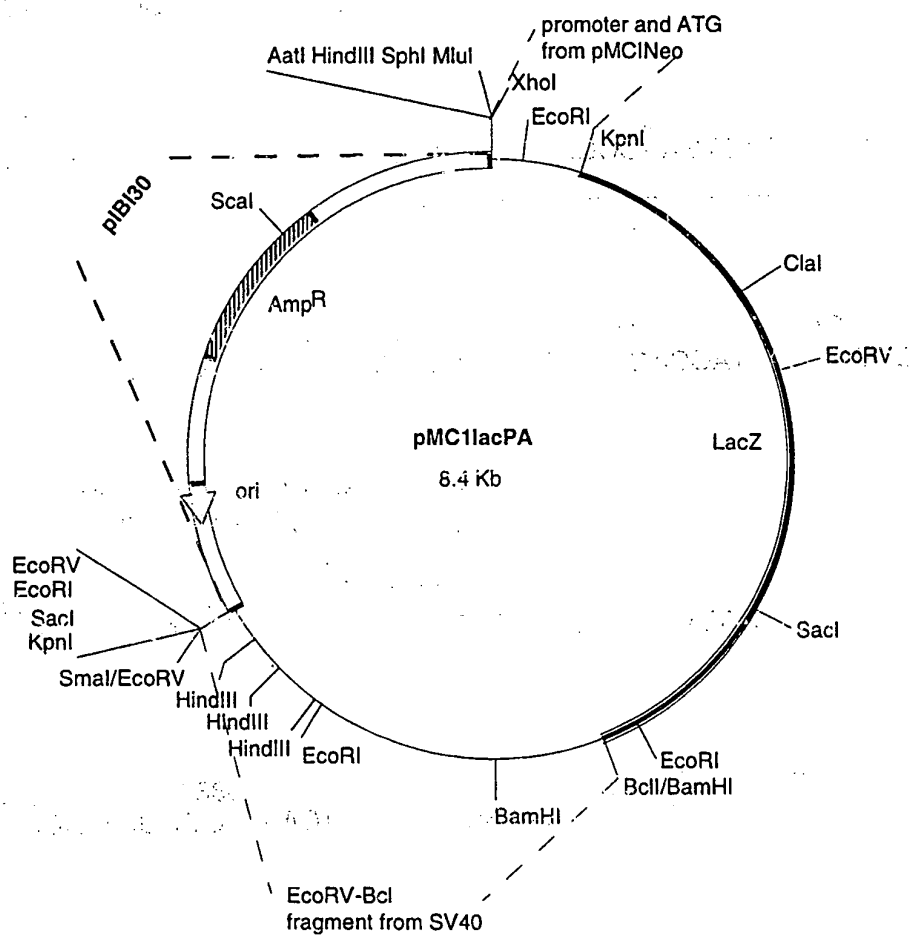
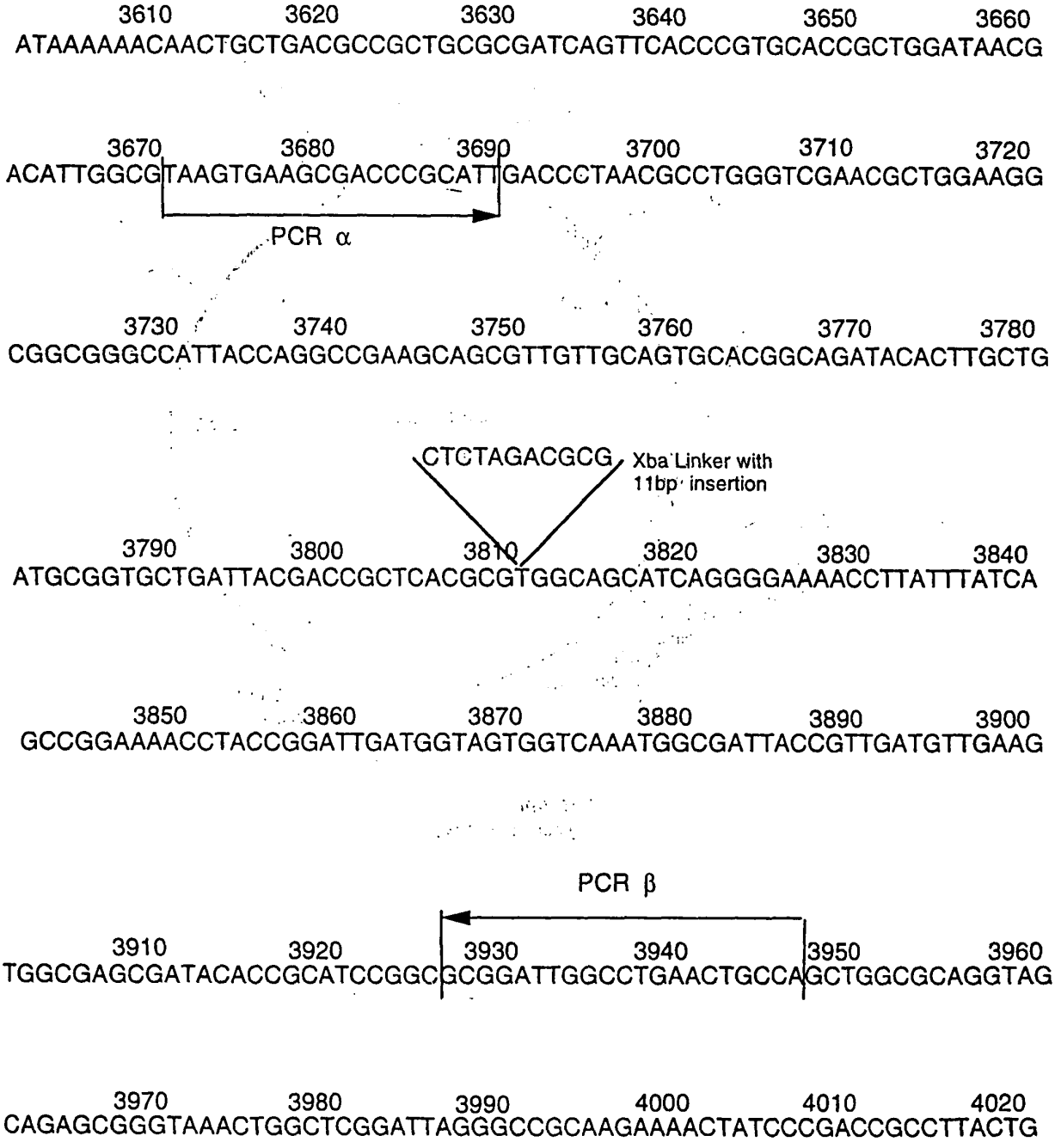

promoter and ATG
from pMClNeo
AatI HindIII SphI MluI
XhoI
EcoRI
KpnI
pIBI30
ScaI
ClaI
AmpR
EcoRV
pMC1lacPA
6.4 Kb
LacZ
ori
EcoRV
EcoRI
SacI
KpnI
SacI
SmaI/EcoRV
HindIII
HindIII
HindIII
EcoRI
EcoRI
BclI/BamHI
BamHI
EcoRV-BclI
fragment from SV40

FIGURE 4

Figure 5


3610 3620 3630 3640 3650 3660
ATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACG
3670 3680 3690 3700 3710 3720
ACATTGGCGTAAGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGG
PCR α
3730 3740 3750 3760 3770 3780
CGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTG
CTCTAGACGCG
Xba Linker with
11bp insertion
3790 3800 3810 3820 3830 3840
ATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCA
3850 3860 3870 3880 3890 3900
GCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAG
PCR β
3910 3920 3930 3940 3950 3960
TGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAG
3970 3980 3990 4000 4010 4020
CAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTG

TEST FOR ALTERATION OF AN INSERTION MUTATION IN THE lacZ (β-GALACTOSIDASE) GENE OF EUKARYOTIC EXPRESSION VECTOR

Experimental Sample	Injected Plasmid, 276-mer DNA and RecA Protein	Number of Injected Surviving Cells	Number of Surviving Cells Scoring Blue	Surviving Cells Scoring Blue (%)
1	pSV-β-gal -276-mer - RecA	168	21	12.5
2	pMC1lacpa - 276-mer - RecA	98	9	9.2
3	pMC1lacXpa - 276-mer - RecA	173	0	0
4	pMC1lacXpa + 276-mer - RecA	103	0	0
5	pMC1lacXpa + 276-mer + RecA	168	6	3.6

Figure 6

Figure 7B

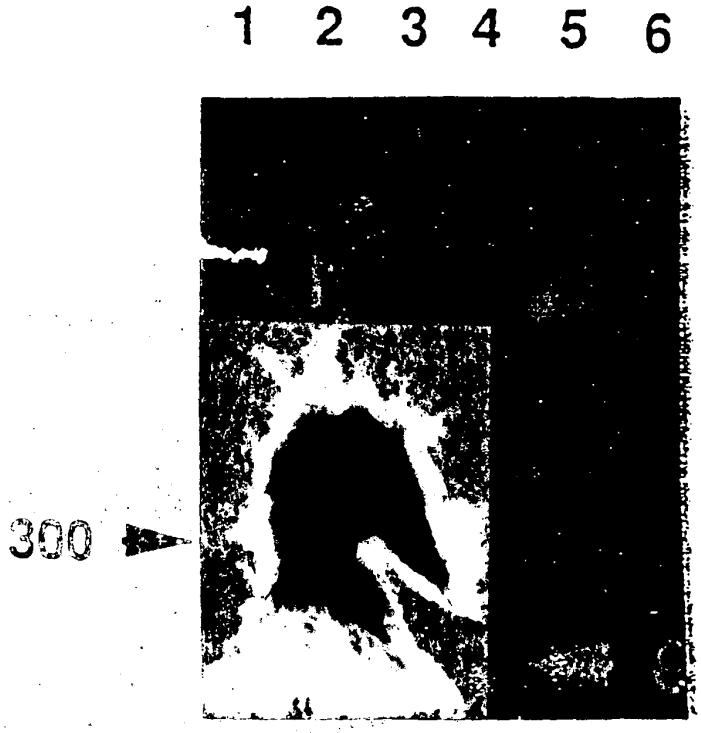


Figure 8A

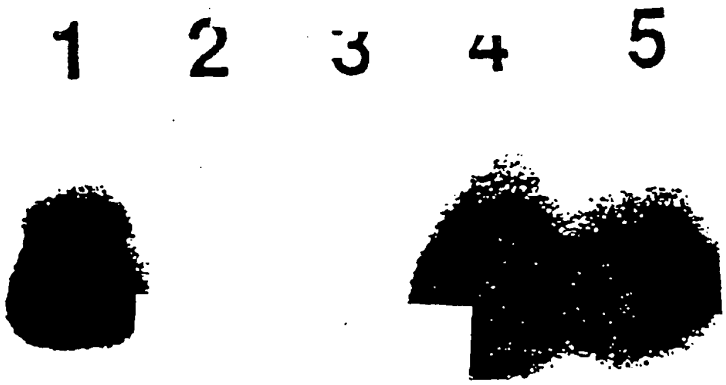
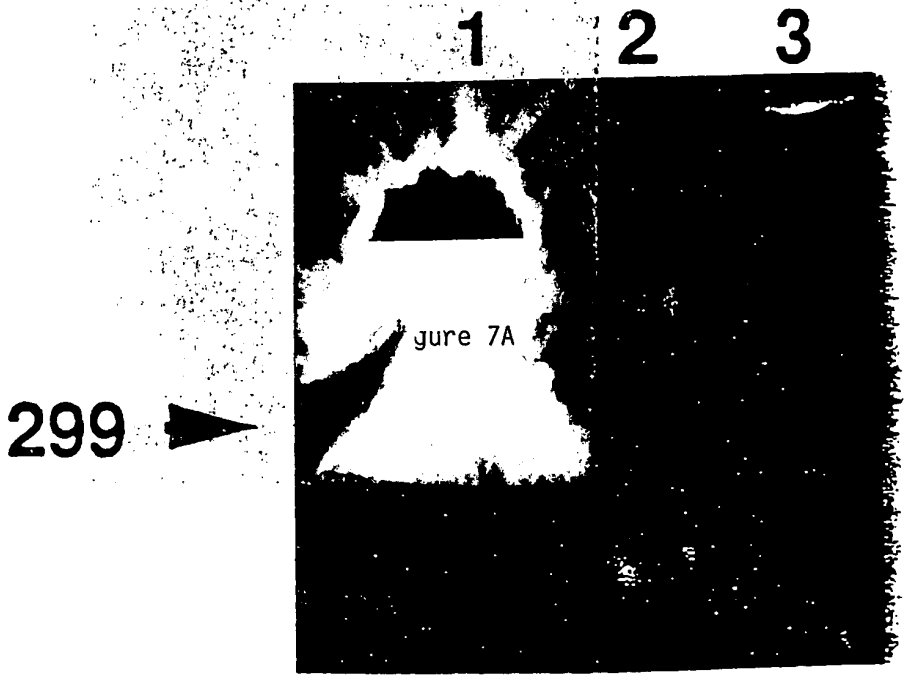


Figure 8B



Fibure 9

Scheme for recombination assay

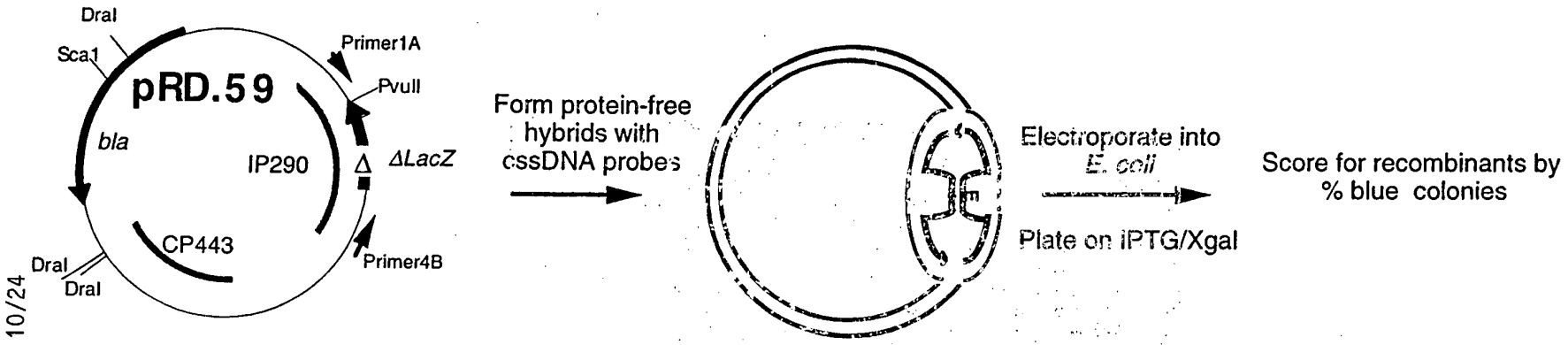


Figure 10

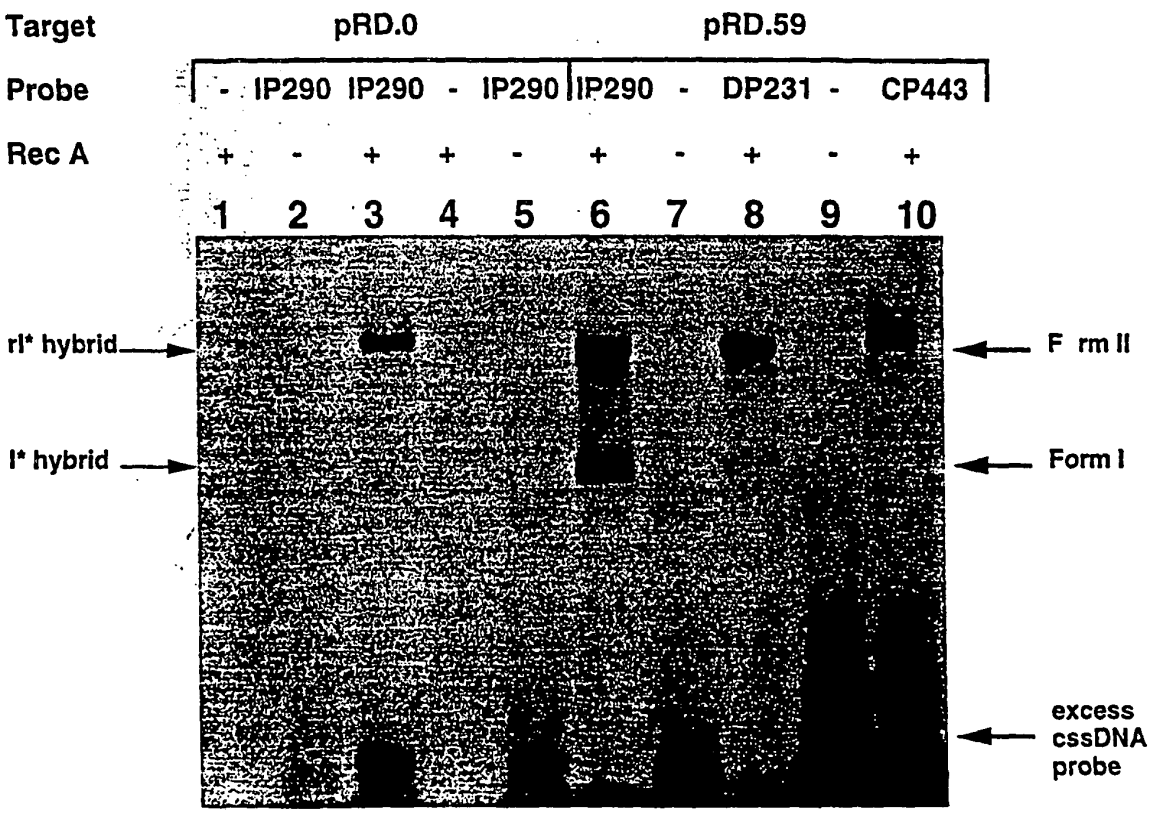

Target
pRD.0
pRD.59
Probe
- IP290 IP290 - IP290 IP290 - DP231 - CP443
Rec A
+ - + + - + - + - +
1 2 3 4 5 6 7 8 9 10
rI* hybrid
I* hybrid
F rm II
Form I
excess cssDNA probe

FIGURE 11

cssDNA Probe:Target Hybrids Enhance Homologous Recombination

Target	Probe	RecA coating	Host	%Recombinant / total colonies
pRD.59	-	+	RecA+	0
	-	+	RecA -	0
pRD.59	IP290	-	RecA+	0
	IP290	-	RecA -	0
pRD.59	IP290	+	RecA+	3
	IP290	+	RecA -	0
pRD.59	DP290	-	RecA+	0
	DP290	-	RecA -	0
pRD.59	DP290	+	RecA+	0
	DP290	+	RecA -	0
pRD.59	CP443	-	RecA+	0
	CP443	-	RecA -	0
pRD.59	CP443	+	RecA+	0
	CP443	+	RecA -	0

Figure 12

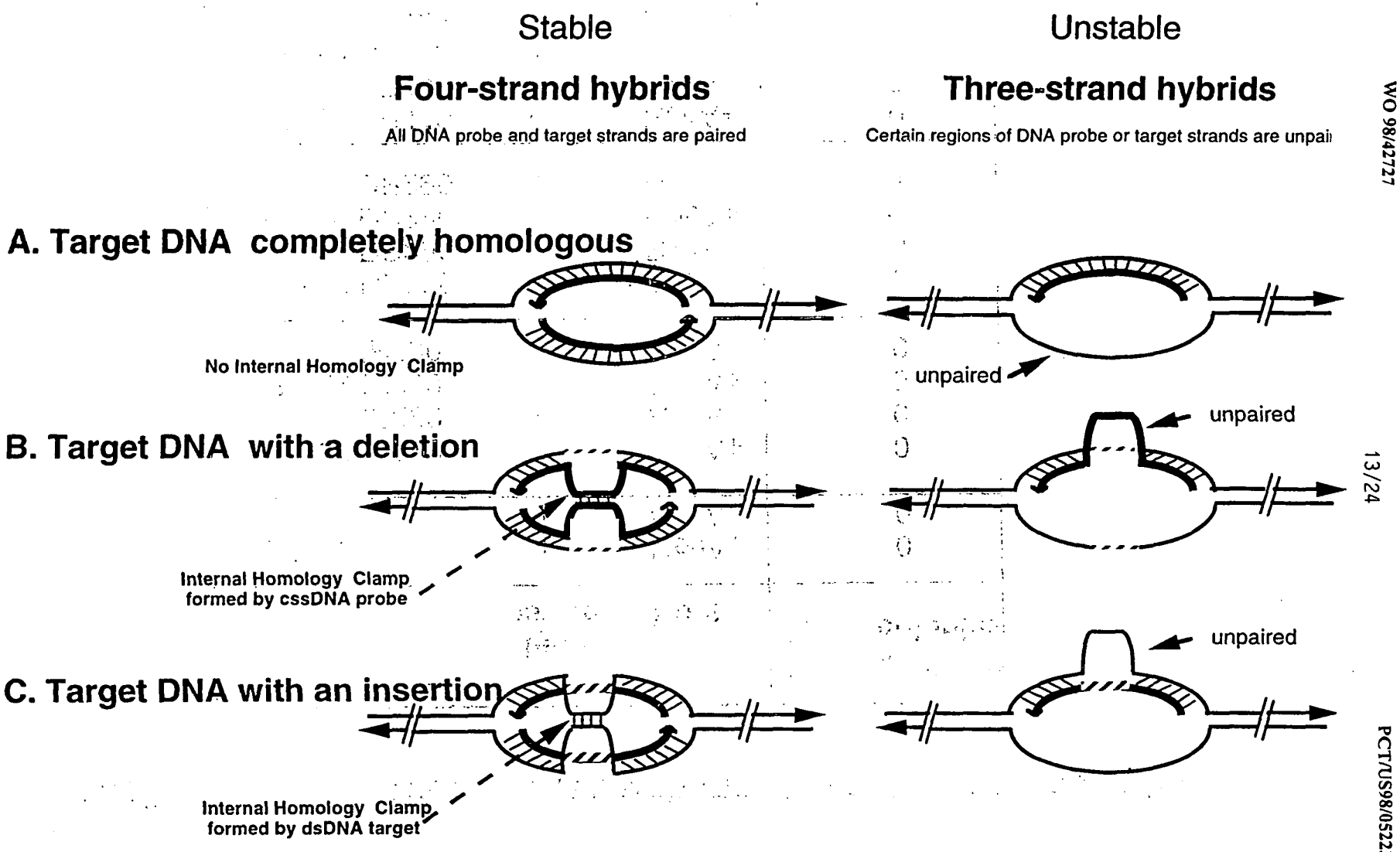

Stable
Unstable
Four-strand hybrids
Three-strand hybrids
All DNA probe and target strands are paired
Certain regions of DNA probe or target strands are unpai
A. Target DNA completely homologous
No Internal Homology Clamp
unpaired
B. Target DNA with a deletion
unpaired
Internal Homology Clamp
formed by cssDNA probe
C. Target DNA with an insertion
unpaired
Internal Homology Clamp
formed by dsDNA target

Figure 13

WO 98/42727 PCT/US98/05223

Figure 14A

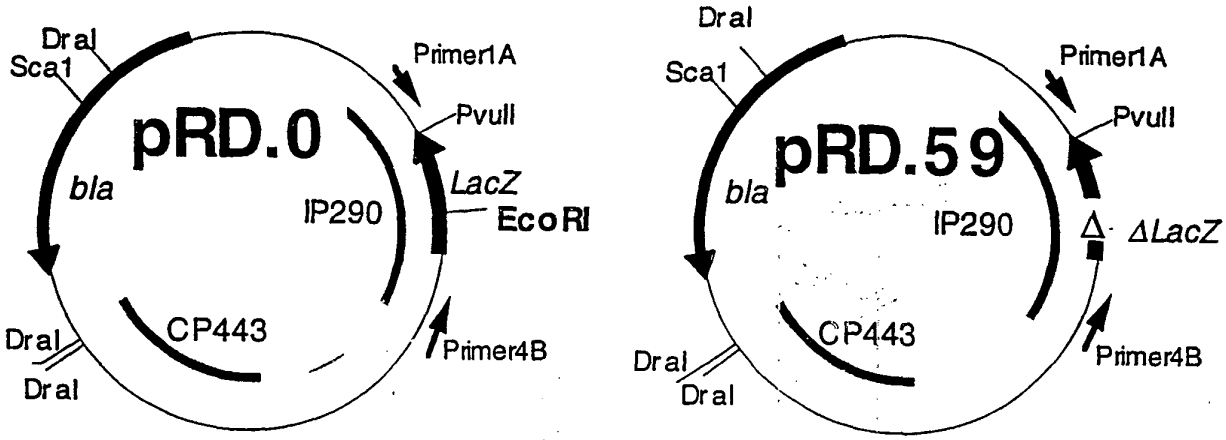
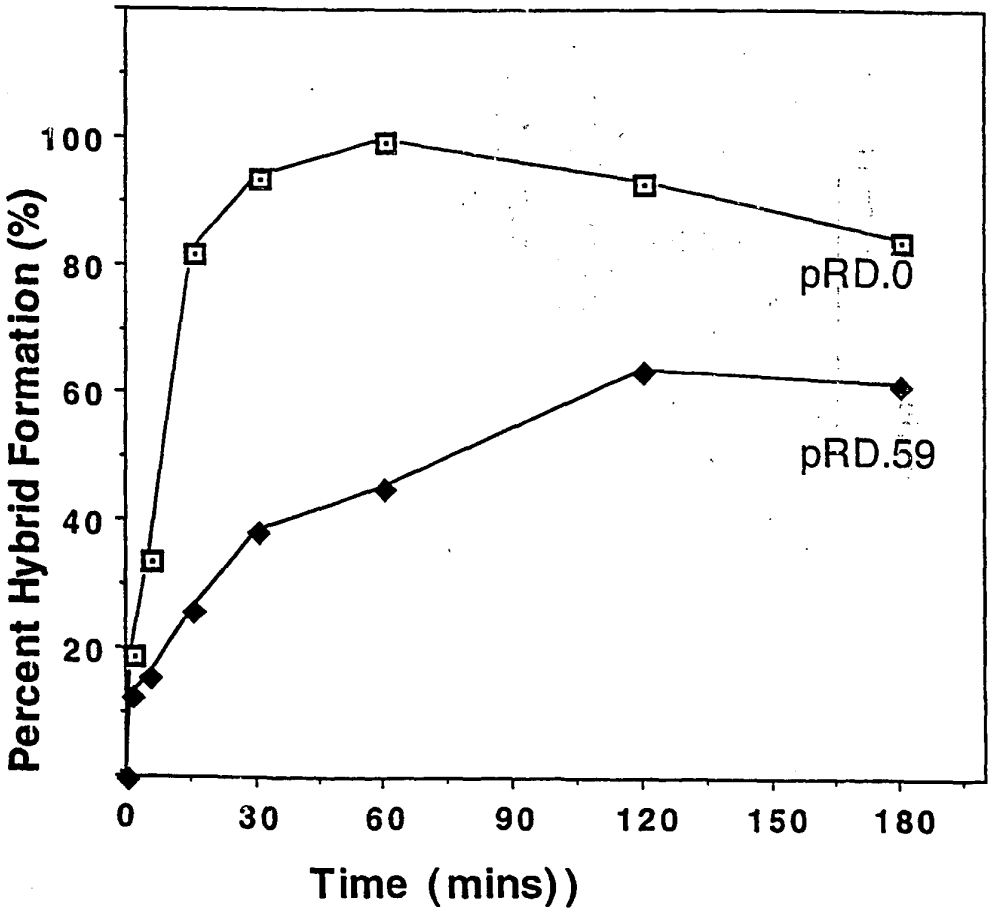


Figure 14B



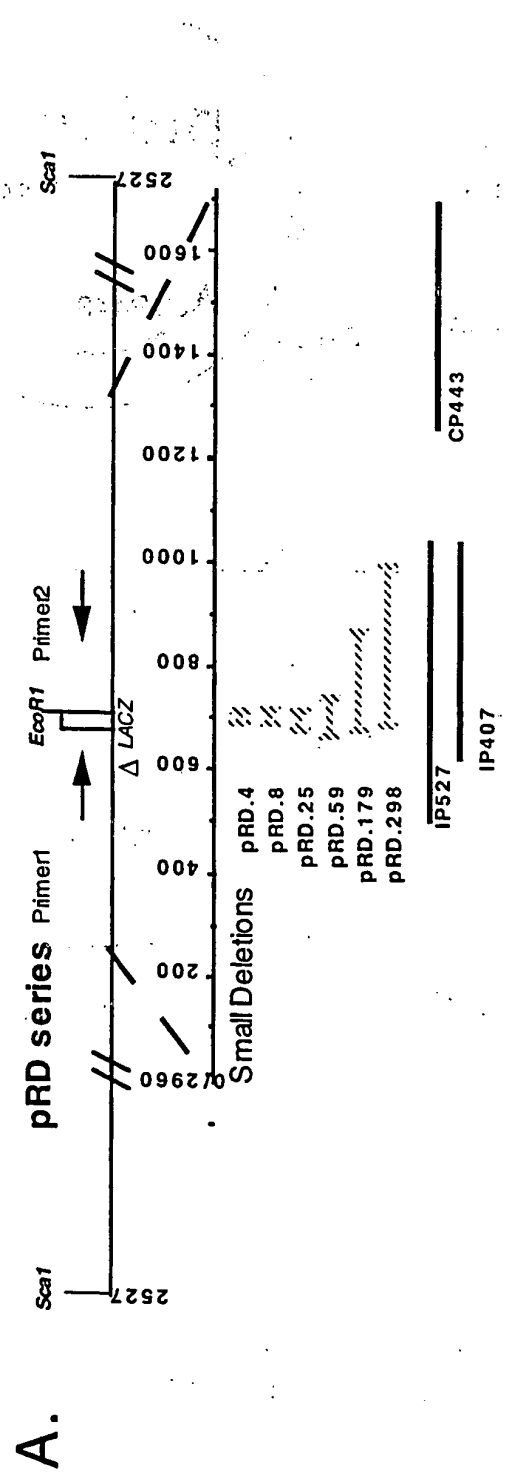

A.
pRD series
Sca1
2527
1600
1400
1200
1000
800
600
400
200
0/2960
Sca1
2527
Primer1
Primer2
EcoR1
Δ LACZ
Small Deletions
pRD.4
pRD.8
pRD.25
pRD.59
pRD.179
pRD.298
CP443
IP527
IP407

Figure 15A

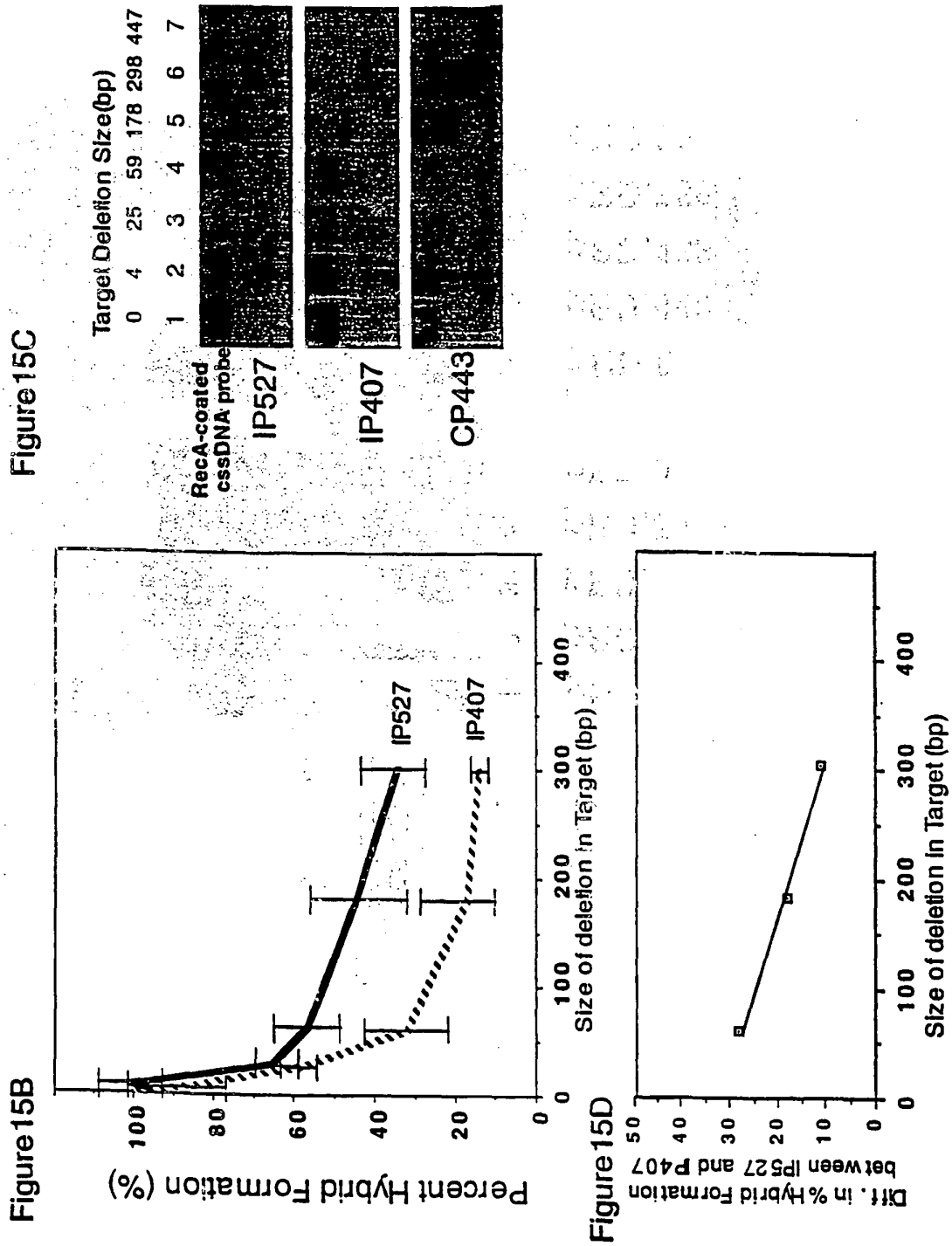
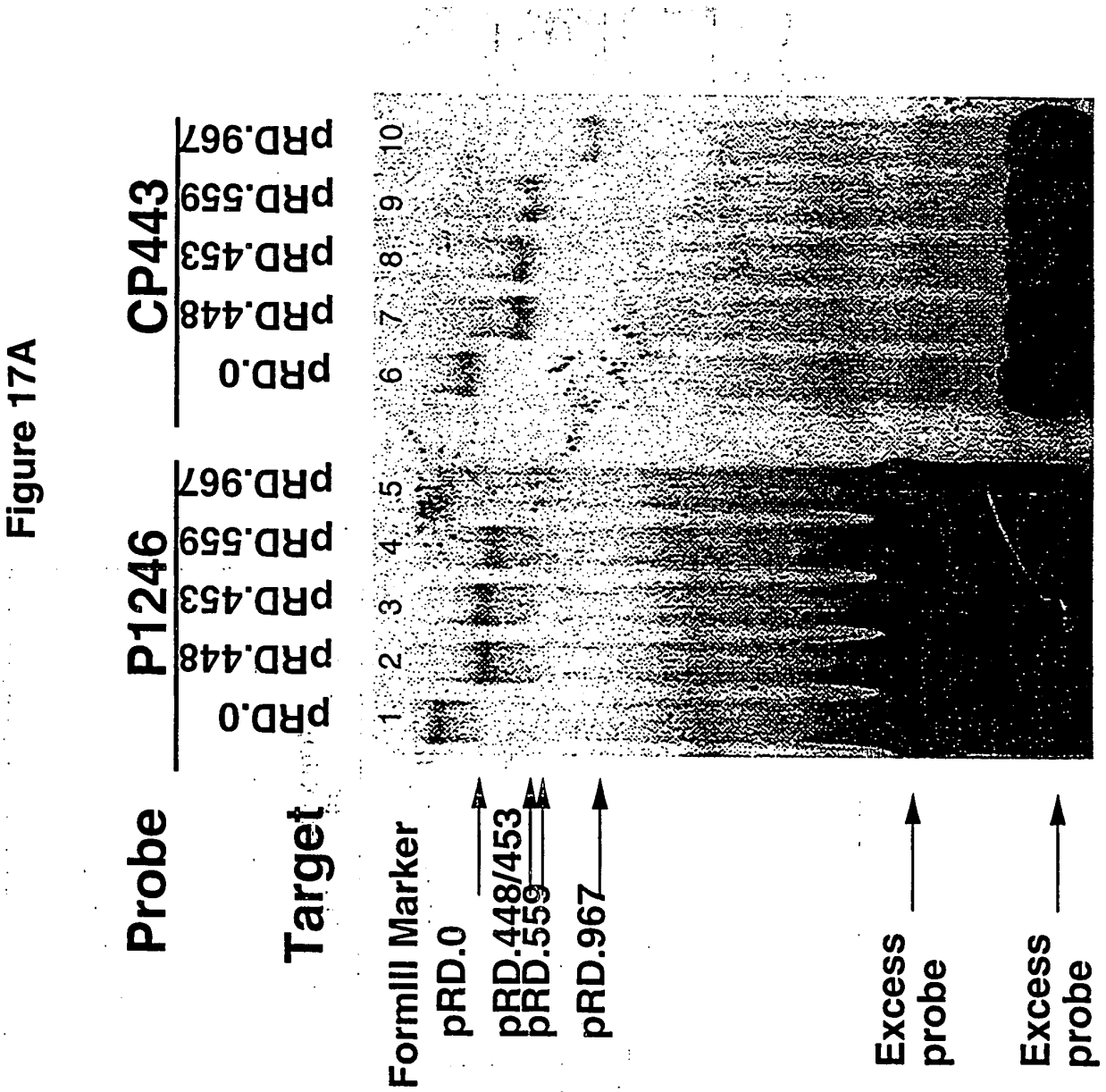


Figure15C
Target Deletion Size(bp)
0 4 25 59 178 298 447
1 2 3 4 5 6 7
RecA-coated cssDNA probe
IP527
IP407
CP443
Figure15B
Percent Hybrid Formation (%)
100
80
60
40
20
0
IP527
IP407
100 200 300 400
Size of deletion in Target (bp)
Figure15D
Diff. in % Hybrid Formation between IP527 and P407
50
40
30
20
10
0
100 200 300 400
Size of deletion in Target (bp)

Figure 17A


Probe
Target
P1246
pRD.0
pRD.448
pRD.453
pRD.559
pRD.967
CP443
pRD.0
pRD.448
pRD.453
pRD.559
pRD.967
1 2 3 4 5 6 7 8 9 10
FormIII Marker
pRD.0
pRD.448/453
pRD.559
pRD.967
Excess probe
Excess probe

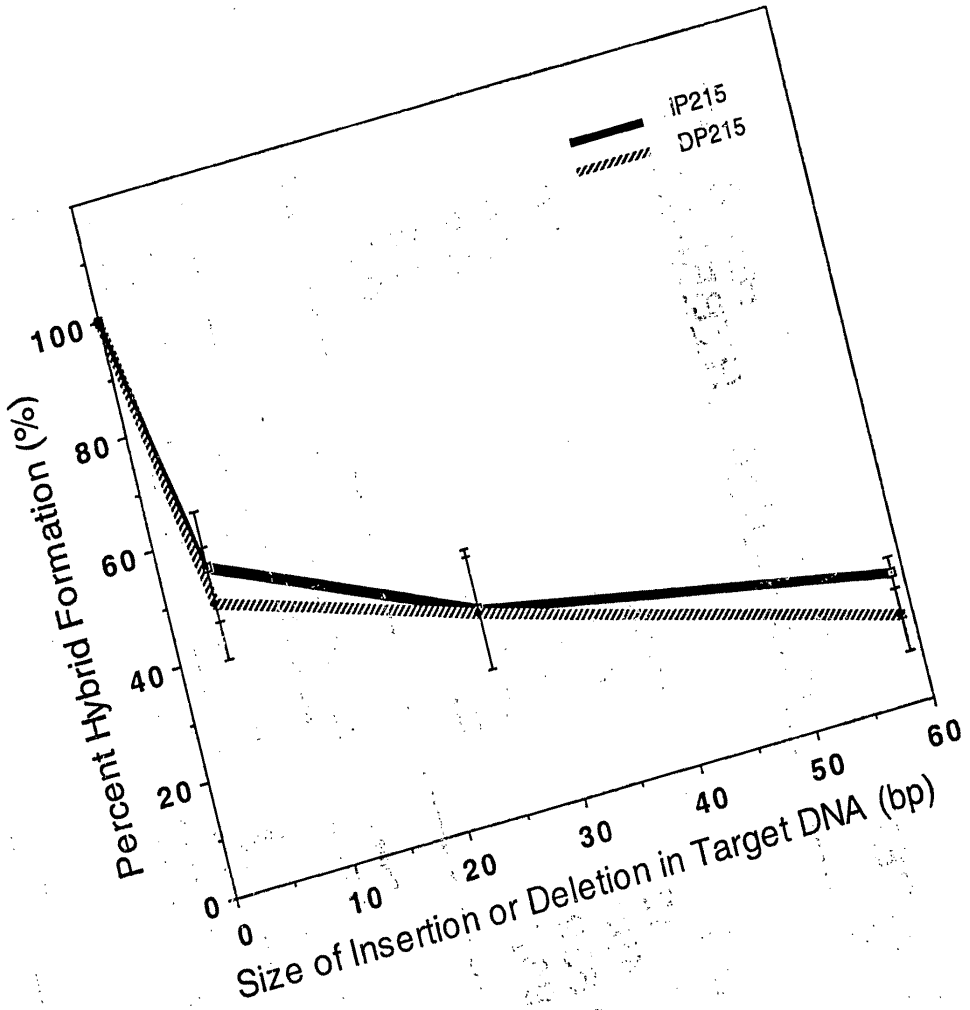

iP215
DP215
Percent Hybrid Formation (%)
100
80
60
40
20
0
0
10
20
30
40
50
60
Size of Insertion or Deletion in Target DNA (bp)

Figure 16

Figure 17B

TARGET DNA	AMT. HOM0L. (bp)	RELATIVE HYBRIDIZATION
pRD.0	1246	1.0
pRD.448	798	0.8
pRD.453	793	0.8
pRD.559	687	0.8
pRD.967	279	0.3

Figure 17C

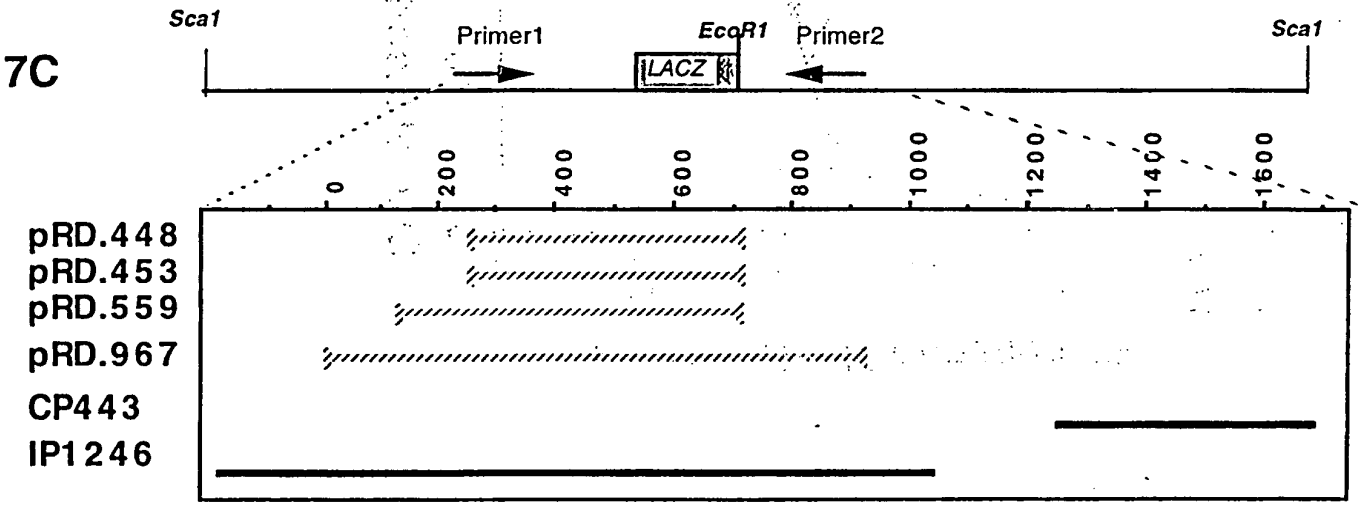


Figure 18

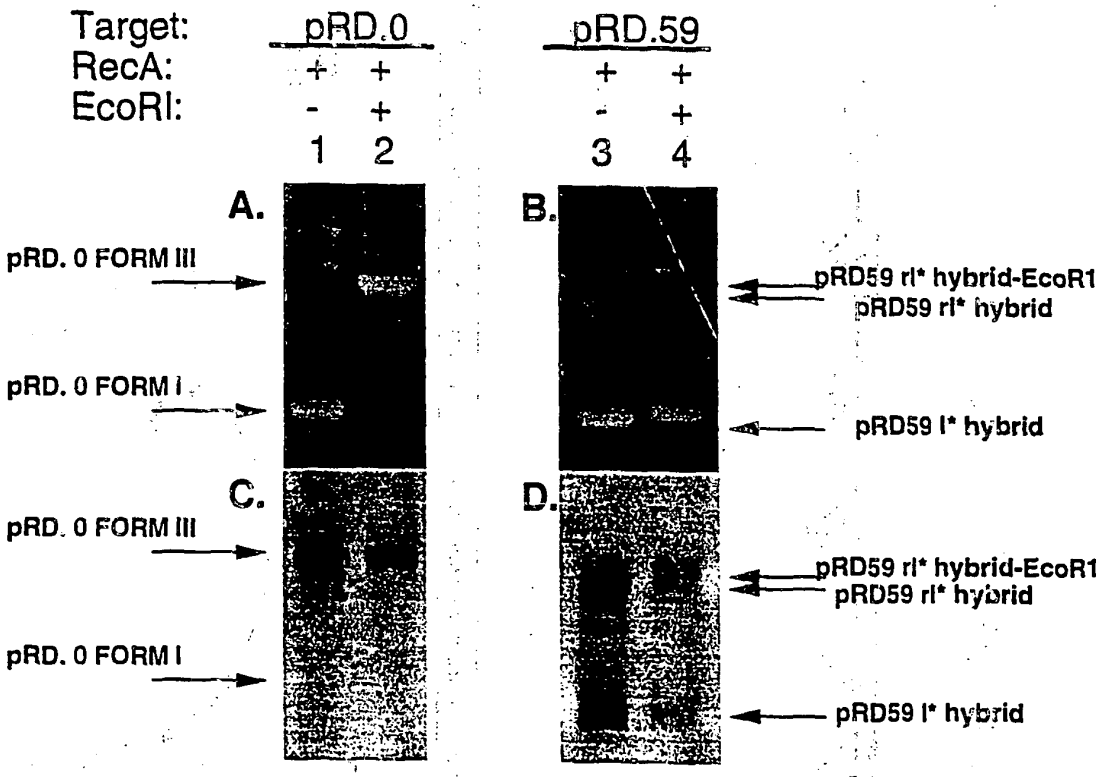
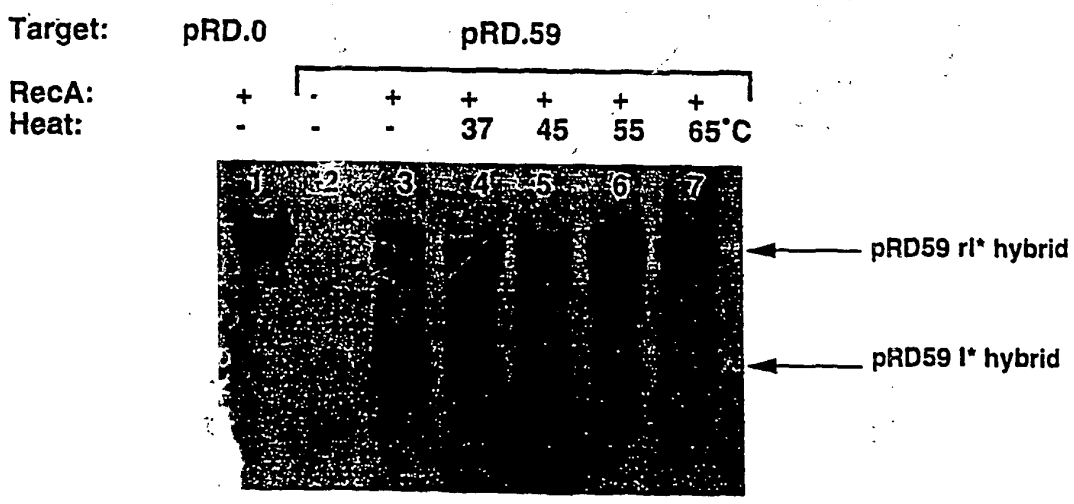


Figure 19



Mouse Ornithine Transcarbamylase Gene

Figure 20A

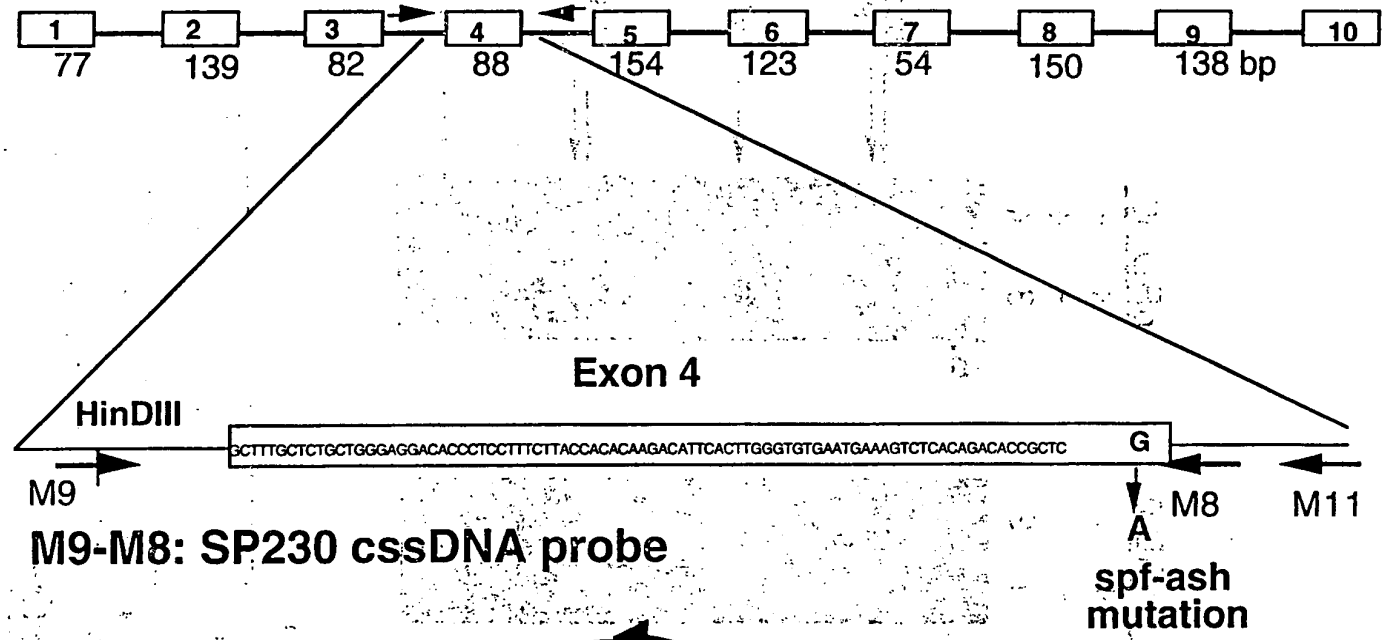


Figure 20B

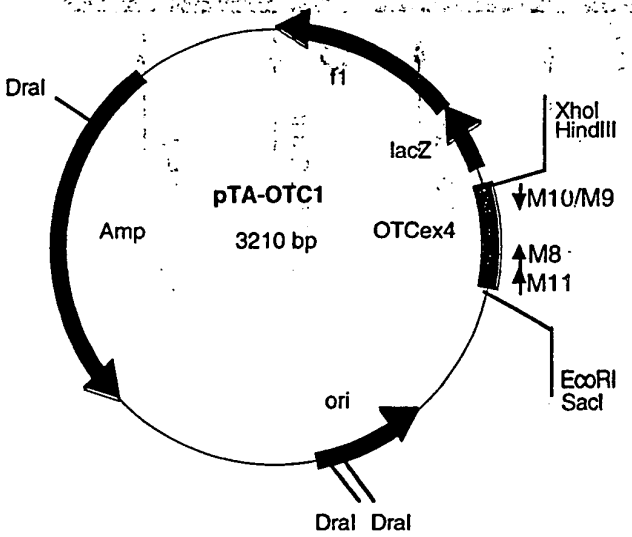


Figure 21 **OTC exon4 DNA sequence determination**

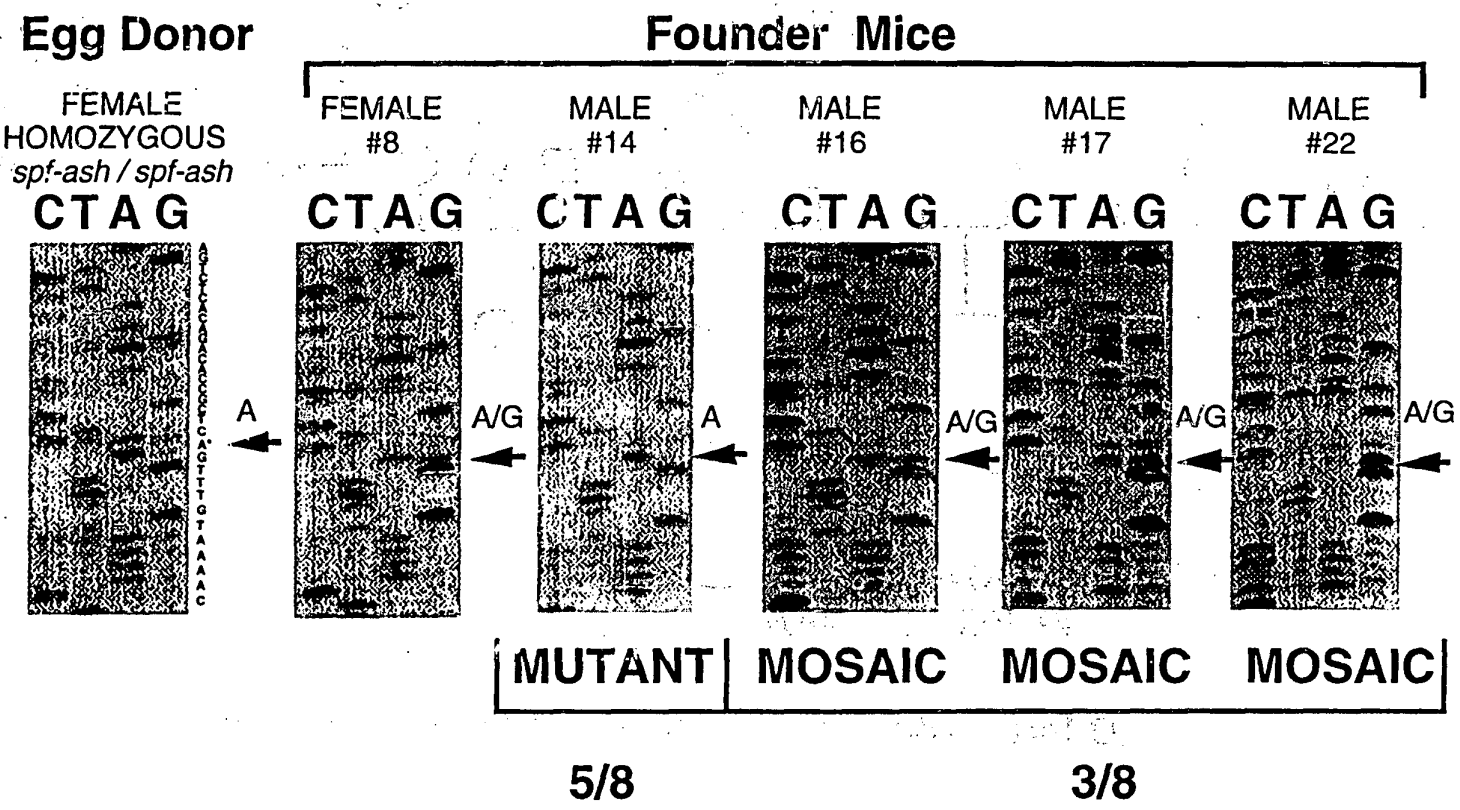


Figure 22 **Germline Transmission of EHR Corrected OTC+ Allele**

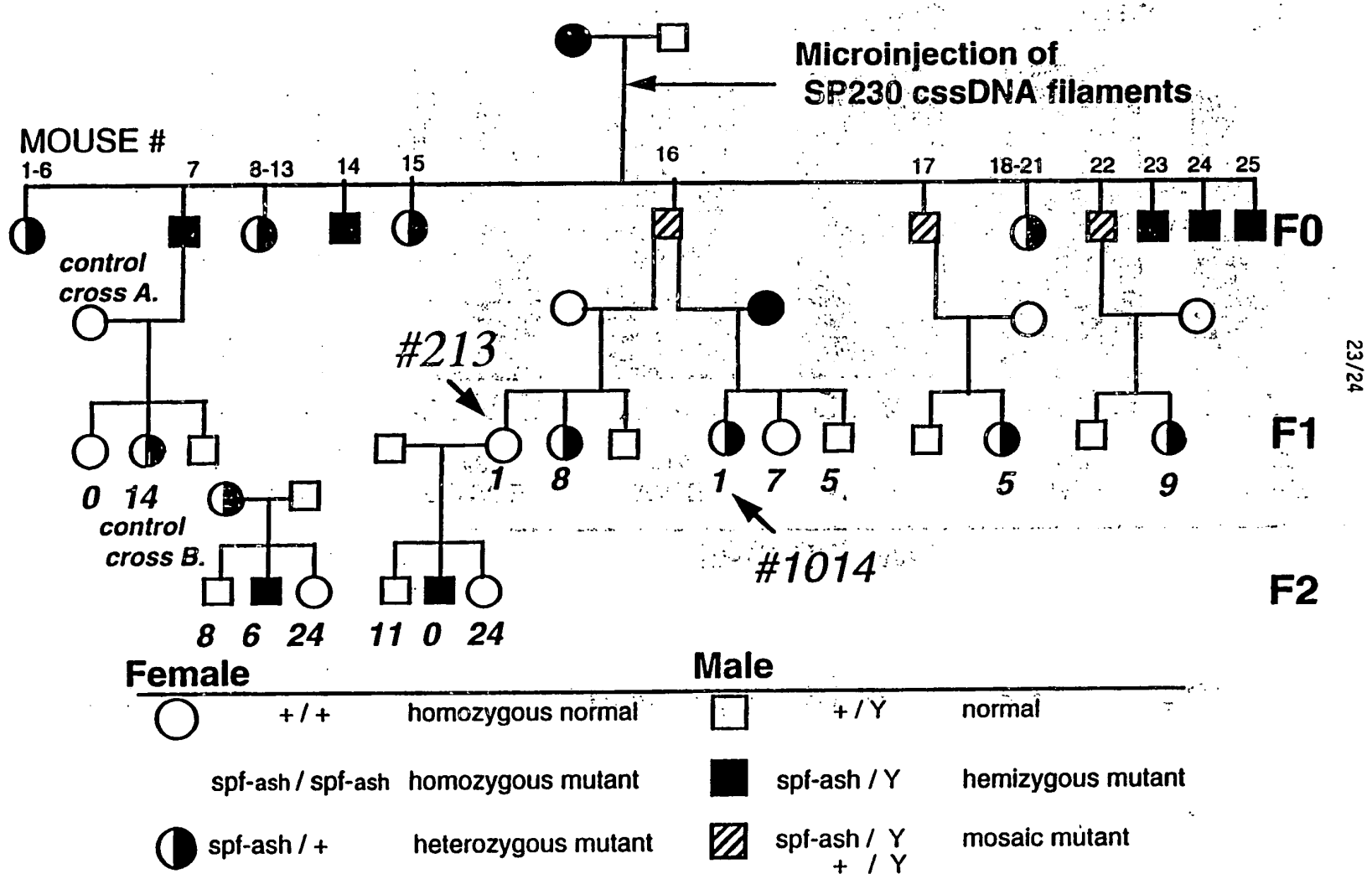


Figure 23 Germline Transmission of Corrected Allele of F0 Male #16
Homozygous spf-ash/spf-ash female parent
Mutant F1 Male Pup1
Mutant F1 Male Pup2
Mutant F1 Female Pup3
Mutant F1 Male Pup5
Mutant F1 Female Pup7
Normal * F1 Female Pup1014